OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: October 1, 1999, 13:54:22 ; Search time 148.69 Seconds
(without alignments)
2793.186 Million cell updates/sec

Title: US-09-026-400-1
Perfect score: 1660
Sequence: 1 attgactagctagttcattc........aaaaaaaaaaaaaaaaaa 1660

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1660	100.0	1660	1	V48147	Nicotianamine amin
2	1146.4	69.1	1910	1	V48148	Nicotianamine amin
c 3	64	3.9	1272	1	X09010	Brn-3a polynucleot
c 4	63.8	3.8	4524	1	X33912	Nucleotide sequenc
c 5	62	3.7	53789	1	V21187	Amycolatopsis medi
c 6	59	3.6	26338	1	V62134	HSV-2 strain SB5 C
c 7	59	3.6	21034	1	V62154	HSV-2 strain SB5 C
c 8	59	3.6	117213	1	V62176	HSV-2 strain SB5 C
c 9	58	3.5	1266	1	X09011	Brn-3a polynucleot
10	56.4	3.4	2061	1	N92408	Sequence encoding
c 11	56.2	3.4	114955	1	X53491	Human adenosine A1
c 12	55.4	3.3	4044	1	Q15154	Plasmid pHAS2-MCP
13	55	3.3	15872	1	T68715	Streptomyces venez
14	54.6	3.3	2186	1	Q99365	S. lividans protea
15	54.4	3.3	2185	1	V84066	Clone P5-4 encodin
16	54.4	3.3	43280	1	T80413	Tylactone synthase
c 17	54.2	3.3	2061	1	N92408	Sequence encoding
18	54.2	3.3	2064	1	Q52638	Streptomyces fradi
19	54.2	3.3	114955	1	X53491	Human adenosine A1
20	53.8	3.2	3993	1	Q15153	Swine herpes virus
21	53.8	3.2	1170	1	Q20217	Sequence of tuf3 g
22	53.8	3.2	29879	1	Q46806	eryA region of S.
23	53.4	3.2	30001	1	T61016	Total DNA sequence
24	53.4	3.2	4257	1	V10362	Infected cell prot
25	53.4	3.2	4257	1	V68520	The nucleotide seq
26	53.4	3.2	30001	1	X05110	S. aureofaciens DN
27	52.6	3.2	390	1	Q21833	Randomising oligon
28	52.6	3.2	390	1	Q36859	PCR primer for 5'
29	52.4	3.2	8169	1	V26609	Actinomadura hibis
30	52.2	3.1	538	1	V44437	Mycobacterium tube
31	52.2	3.1	538	1	V64546	M. tuberculosis im
32	51.8	3.1	15672	1	Q10613	Rianodin receptor
33	51.8	3.1	24379	1	T93095	Streptomyces freno
34	51.8	3.1	24379	1	V25925	Streptomyces roseo
35	51.2	3.1	36519	1	V22141	Chimpanzee adenovi
36	51.1	3.1	2752	1	V11736	Barley alpha-gluco
37	50.4	3.0	758	1	V99751	M. luteus RP-facto
38	50.2	3.0	1269	1	T33617	DxrA gene encoding
c 39	50.2	3.0	8051	1	T72685	Sugar biosynthesis
40	50.2	3.0	8160	1	X25774	S.erythraea erythr
41	50	3.0	1748	1	T11744	Human kynurenine a
42	49.6	3.0	1181	1	Q85933	cDNA encoding Dac
43	49.6	3.0	1671	1	T85645	DNA encoding enzym

c 44 49.6 3.0 2312 1 X16152 Chicken Sox1 cDNA.
45 49.2 3.0 1076 1 X13772 Enterococcus faeca

ALIGNMENTS

RESULT 1
V48147
ID   V48147 standard; cDNA to mRNA; 1660 BP.
AC   V48147;
DT   27-OCT-1998 (first entry)
DE   Nicotianamine aminotransferase 49564.15 molecular weight protein, gene.
KW   ds; nicotianamine aminotransferase; plant; iron absorption;
KW   iron deficiency chlorosis.
OS   Gramineae sp.
FH   Key   Location/Qualifiers
FT   CDS   62..1447
FT   /*tag= a
FT   /product= "Nicotianamine aminotransferase"
PN   EP-860499-A2.
PD   26-AUG-1998.
PF   19-FEB-1998; 102891.
PR   21-FEB-1997; JP-037499.
PA   (SUMO) SUMITOMO CHEM CO LTD.
PI   Mori S, Nakanishi H, Takahashi M;
DR   WPI; 98-439341/38.
DR   P-PSDB; W61642.
PT   New nicotianamine aminotransferase protein and DNA - useful for
PT   enhancing iron absorption of plant cells
PS   Claim 4; Page 12-13; 17pp; English.
CC   The nicotianamine aminotransferase can be used in a plasmid to transform
CC   plant cells to produce cells with enhanced iron absorption, and it is
CC   implied (though not stated) that plants with improved resistance to iron
CC   deficiency chlorosis in calcareous soils can be regenerated from the
CC   transformed cells. The gene fragment can be used to detect, amplify
CC   and/or isolate nicotianamine aminotransferase genes.
SQ   Sequence 1660 BP; 423 A; 442 C; 430 G; 365 T;

Query Match   100.0%; Score 1660; DB 1; Length 1660;
Best Local Similarity   100.0%; Pred. No.0;
Matches 1660; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 attgactagctagttcattcctgccacactgctagtactcctcgtttcctcgtggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATTGACTAGCTAGTTCATTCCTGCCACACTGCTAGTACTCCTCGTTTCCTCGTGGC 60

Qy   61 aatggtacaccagagacaacggccacggccgagcgccgccgccaacggcaagag 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AATGGTACACCAGAGACAACGGCCACGGCCGAGCGCCGCCGCCAACGGCAAGAG 120

Qy   121 caacgggcaccgccgccgccgcaacggcaagagcaacggcaacggcggcggccggt 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAACGGGCACCGCCGCCGCCGCAACGGCAAGAGCAACGGCAACGGCGGCGGCCGGT 180

Qy   181 ggagtggaattcgcccgggccaagacgacggcatcctggcgacgacggggcaagacag 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGAGTGGAATTCGCCCGGGCCAAGACGACGGCATCCTGGCGACGACGGGGCAAGAACAG 240

Qy   241 catccgggcgatacggtacaagatcagcgagcgtgagagaggagcgggccgccgt 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CATCCGGGCGATACGGTACAAGATCAGCGAGCGTGAGAGAGGAGCGGGCCGCCGT 300

Qy   301 gctgccgctggccacgtgccacggtgaccgtgttccggccttccgcacggccgtcgaggc 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GCTGCCCGCTGGCCACGTGCCACGGTGACCGTGTTCCGGCCTTCCGCACGGCCGTCGAGGC 360

Qy   361 cgaagacgccgtcgccgccgccgcgccgctggccgtcacggccagttcaactgctacgcccgcggcgt 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CGAAGAGCCGCCGTCGCCGCCCGCCGCCTGCCGCCAGTTCAACTGCTACGCCCGCGGCGT 420

Qy   421 cggcctcccgcccgcgcacgaagcgccgtagcagagcacttgtcacaggcgtgtcacctacaa 480

```
Db  1501 ATCCCCATTTATATATTTCAATAAAACATATTGTTGTTGTTGTTGAGCTGTCCAAG 1560
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1561 tcatgtactctactttttgatgtattggcctcattgcttgcttgcatcaattcaataaaa 1620
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 TCATGTACTCCACTTTTTGATGTATTTGGCCTCATTGCCTTGCTTGCATCAATTTCAATAAAAA 1620

Qy  1621 tggttgtgtacaccaaaaaaaaaaaaaaaaaaaa 1660
         ||||||||||||||||||||||||||||||||||
Db  1621 TGGTTGTGTACACCAAAAAAAAAAAAAAAAAAAAA 1660
```

RESULT   2

```
V48148
ID     V48148 standard; cDNA to mRNA; 1910 BP.
AC     V48148;
DT     27-OCT-1998 (first entry)
DE     Nicotianamine aminotransferase 58148.62 molecular weight protein, gene.
KW     ds; nicotianamine aminotransferase; plant; iron absorption;
KW     iron deficiency chlorosis.
OS     Gramineae sp.
FH     Key             Location/Qualifiers
FT     CDS             76..1731
FT                     /*tag= a
FT                     /product= "Nicotianamine transferase"
PN     EP-860499-A2.
PD     26-AUG-1998.
PF     19-FEB-1998; 102891.
PR     21-FEB-1997; JP-037499.
PA     (SUMO ) SUMITOMO CHEM CO LTD.
PI     Mori S, Nakanishi H, Takahashi M;
DR     WPI; 98-439341/38.
DR     P-PSDB; W61643.
PT     New nicotianamine aminotransferase protein and DNA - useful for
PT     enhancing iron absorption of plant cells
PS     Claim 4; Page 14-15; 17pp; English.
CC     The nicotianamine aminotransferase can be used in a plasmid to transform
CC     plant cells to produce cells with enhanced iron absorption, and it is
CC     implied [though not stated] that plants with improved resistance to iron
CC     deficiency chlorosis in calcareous soils can be regenerated from the
CC     transformed cells. The gene fragment can be used to detect, amplify
CC     and/or isolate nicotianamine aminotransferase genes.
SQ     Sequence  1910 BP;   462 A;   534 C;   546 G;   368 T;
```

```
Query Match          69.1%;   Score 1146.4;   DB 1;   Length 1910;
Best Local Similarity  83.8%;   Pred. No. 1.9e-211;
Matches 1333;  Conservative    0;  Mismatches 226;  Indels   31;  Gaps    2;
```

```
Qy   68 caccaggagcaacggccacgcgaggccgcgccggccgccgccaacggcaagagcaacggg 127
         |||||| |||||||| |||| |||||||  ||| ||| |||||||||||||||||| |||
Db  337 CGCGAGGAGCAACGGCCATGCCATGCTCGAGGCCGCCGACGGCGCGAGAGCAACGAGCAATGCC 396

Qy  128 cacgcgcgccgccgaacggcaagagcaacgcggcgcacgcggcgcggc---------gcggc 175
         |||| ||| ||| ||||||||||||||||| || |||||||| |||
Db  397 GAGGACTCGCCGCCGGCGGAGCAACGGCGCCGGCCGCCGGCGCAGGAGGAGGAG 456

Qy  175 --ggcggtggagtggaatttcgccgggtcgaggatcagccgtacagatcagccgagcgtacaga 232
           |||||| ||||||||||||||||||||||||||||| ||| ||||||||| ||||| |||
Db  457 GAGGCGGGTGGAGTGGAATTTCGCGGCGAGTCTACTAAAGGAATCATC 516

Qy  233 aagaacagcatccggcgtacacgtacagatcagccgagctagacggatcagccgagcgtgcaggagaagcgg 292
         |||  ||| | |||| | | ||||| ||||||||||| ||||| |||||  |||||||| |||||||| |||||
Db  517 AACATGAGCATCCGGGCGATACGGGCGATACGGTACAGATCAGCCGCTGCGCAGAGGGGCCG 576

Qy  293 cggcccgtgctgccgctggccaccgagggaagatccatgacga 352
         |||||| ||||| || |||| || |||||||| ||| |||  ||
Db  577 CGGCCCGTGCTGCTGCCGCTGGCCACCGGGACCGCGTGTTCCCGCCCGTTCCCACGCGCC 636

Qy  353 gtcgagccgaagacgacgccgtcccgcgcgtgcgtgcaaccggcaactgctacgcc 412
         ||||| |||||||| | || ||| ||||| |||| ||| | |||| |||||||  || ||
Db  637 GTCGAGGCGGCCGAGGACGACGCCGCCCGCCCCGTGCCGCCGCCCAGTTCAACTCTCTACCCC 696

Qy  413 gccggcgtcggcctcccccgcgcacgaagcgcctcccccgccgtagcagcagcgccgacacttgtcacacgggggctgt 472
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  421 CGGCTCCCGCGCCCACGAGCGCGTAGCAGAGCAGCGCGCGTGCCTGCCTGCCTGCCCTACAA 480
         |||||||||||||||||||||||||||||||||||||||||||||
Qy  481 gctatcggccgacgacgtcttcctcaccgcgcggaactcaggcgatcgaagtcataat 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GCTATCGGCCGACGAGCGTCTTCCTCACCGCGCGGAACTCAGGCGATCGAAGTCATAAT 540
Qy  541 ccggtgctggccagactgcgcggcgccaacatactgcttcccggccaggctatccaaa 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CCCGGTGCTGGCCAGAGCTGCGCGGCGCGCCAACATACTGCTTCCCGGCCAGGCTATCCAAA 600
Qy  601 ttacgagggcgcagcggcattcaacaagctggagtccggcacttcgacctcatccccga 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TTACGAGGGCGGAGCGGCCATTCAACAAGCTGGAGTCCGGCCACTCGACCTCATCCCCGA 660
Qy  661 caagggctgggagatcgacatcgactcgctgtgaatccatcgccgacaagaacaccaccgc 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 CAAGGGCGTGGGAGATCGACATCGACATCGAATCCATCGCCGACAAGAACAACCACCGC 720
Qy  721 gatggtcatcataaacccaacaatccgtgcgcagcgttactcctacgaccatctgcc 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GATGGTCATCATCAAAACAATCCGTGCGCAGCGTTTACTCCTACGACCATCTGCC 780
Qy  781 caaggtcgcggaggtggcaaggaagctccggaatattggtgatcgctgacgaggtttacgg 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CAAGGTCGCGGAGGTGGCAAGGAAGCTCGGAATATTGGTGATGCTGACGAGGTTTACGG 840
Qy  841 caaactggttctcggcagcgcccgttatccgatgggcgtcttgggcacattgcccc 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 CAAACTGGTTCTGGGCAGCGCCGCCCGTTTATCCGATGGGCGTCTTTGGGCACATTGCCCC 900
Qy  901 ggtcttgtccattggatcctctgtccaagtcgtggatagtgcgacttggatg 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 GGTCTTGTCCATTGGATCTCGTCTGTCCAAGTCGTGGATGGCGACTTGGATG 960
Qy  961 ggtggcggtgtacgaccccacaaagatttagagaaactaagatcctctacgtctattac 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 GGTGGCGGTGTACGAGCCCCACAACGATTTTAGAGAAAACTAAGATCCTCTACGTCTATAC 1020
Qy  1021 gaattacctaatgtctcaacgaccagcaaccttcgttcaggagagctcttcctaaaat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 GAATTACCTTAATGTCTCAACGACCCAGCAACCTTCGTTCAGGAGACTCTTCCTAAAAT 1080
Qy  1081 tcttgagaacaacaaagcagatttcttaagaggattattggtctactaaaggaatcatc 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 TCTTGAGAACAACAAAGCCACAAGCCAGATTTCTTTAAGAGGATTATTGGCTACTAAAGGAATCATC 1140
Qy  1141 agagatatgttagggaaataaggaaataaagtaaagagaaacaaatatattacgtgtcctcacaagccaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 AGAGATATGTTATAGGGAAATAAAGGAAATAAAGAAAACAAATATATTACCGTGTCCTCACAAGCCAGA 1200
Qy  1201 aggatcgatgtttgtaatggtcaaactaaactctttggaggagatccatgacga 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AGGATCGATGTTTGTAATGTCAAACTAAACTTACATCTTTGGAGGAGATCCATGACGA 1260
Qy  1261 catagattttgctgcaagctcgcaaaggaagaaatcagtaattttatgtccaggggtgt 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CATAGATTTTTGCTGCAAGCTCGCAAAGGAAGAAATCAGTAATTTTATGTCCAGGGGAGTGT 1320
Qy  1321 tcttggaatgaaaattgggtccgtattactttgcctggttccatcttcttcttcaaga 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 TCTTGGAATGAAAATTGGGTCCGTATTACTTTGCCTGGTTCCATCTTCTTCTTCAAGA 1380
Qy  1381 tggactgaaagggtcaaatcattctgtcaaaggaacaagaagaagaattctataaatgg 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 TGGACTGAAAGGGTCAAATCAATATCATTCTGTCAAAGGAACAAGAAGAAGAATTCTATAAATGG 1440
Qy  1441 ttgttagttgtacacaccccagttgtacatcgtgaagctgtaaatcattctagtt 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 TTGTTAGTTGTACACACCCCAGTTGTACATCGTGAAGCTGTAAATCATTTCTAGTT 1500
Qy  1501 atccccattttatatttcaataaacatattgtaatggttgtgttgtagctgtccaag 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1761  TCCCAGTATCCCCATCTATATCTTTCAATAAAATGGAACTTTAGTTCTCTATGAATAGA  1820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1553  tgtccaagtcatgtactctacttttgatgtattggcctcattgctcatcaatttc  1612

Db   1821  AGTCAAACACTCCTGAATATGTTCTGGTTGTTGTGGCCTGGACGAAACATAGTGAATGT  1880
           | || || |||  ||||  || | |  | ||| | ||| || |  ||| ||| |
Qy   1613  aataaaatggttgtgtaccaccaaaaaaa  1642

Db   1881  TATGTTAGTGAAGTTAAAAAAAAAAAAAA  1910
```

```
RESULT   3
X09010/c
ID   X09010 standard; cDNA; 1272 BP.
AC   X09010;
DT   14-JUN-1999  (first entry)
DE   Brn-3a polynucleotide.
KW   Brn-3a; Bcl-2; neurons; neuronal cells; apoptosis; cell death; CNS;
KW   PNS; central nervous system; parasympathetic nervous system;
KW   development; injury; neurotrophic; nerve growth factor; NGF;
KW   ciliary neurotrophic factor; CNTF; brain-derived neurotrophic factor;
KW   BNTF; neurotrophin; NT-3; NT-4; NT-5; neurodegenerative disease;
KW   familial dysautonomia; infantile muscular dystrophy;
KW   Parkinson's disease; Alzheimer's disease; ss.
OS   Homo sapiens.
FH   Key             Location/Qualifiers
FT   tag             1..1272
FT                   /*tag= a
FT                   /product= Brn_3a_polypeptide
PN   WO9905272-A1.
PD   04-FEB-1999.
PF   27-JUL-1998;  G02228.
PR   10-DEC-1997;  US-988476.
PR   25-JUL-1997;  GB-015823.
PA   (UNLO ) UNIV COLLEGE LONDON.
PI   Latchman DS, Smith MD;
DR   WPI; 99-142928/12.
DR   P-PSDB; W96262.
PT   New polypeptide comprising transcription factor Brn-3a, or its
PT   derivative - useful for treating nervous system diseases, preventing
PT   cellular apoptosis and increasing nerve regeneration following
PT   neuronal damage
PS   Disclosure; Page 61-62; 68pp; English.
CC   Over expression of transcription factor Brn-3a can protect neuronal
CC   cells from apoptosis. Brn-3a also specifically activates expression
CC   of the Bcl-2 gene in neuronal cells and this activation is mediated
CC   via a Brn-3a response element in the 5' regulatory region of the
CC   Bcl-2 gene. Both the anti-apoptopic effect of Brn-3a and its ability
CC   to activate expression of Bcl-2 are mediated by the N-terminal
CC   domain of Brn-3a. Members of the Bcl-2 family perform critical roles
CC   in the regulation of selective apoptosis during development of the
CC   nervous system. The stimulation of Bcl-2 expression by Brn-3a in a
CC   neuron specific manner and consequent protection of neuronal cells
CC   from apoptosis suggests that Brn-3a may co-ordinate some aspects of
CC   neuronal reorganisation during development or following injury. The
CC   elevation of Brn-3a expression by either pharmacological means
CC   (compositions comprising one other therapeutic polypeptide e.g.
CC   neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic
CC   factor (CNTF), brain-derived neurotrophic factor (BNTF), and
CC   neurotrophins NT-3 and NT-4/5) or gene therapy may represent a
CC   method for treating human diseases associated with excessive
CC   neuronal cell death and/or lack of nerve regeneration, especially
CC   neurodegenerative diseases such as familial dysautonomia and
CC   infantile muscular dystrophy, and Parkinson's and Alzheimer's
CC   disease.
SQ   Sequence   1272 BP;   202 A;   468 C;   452 G;   150 T;

Query Match             3.9%;  Score 64;  DB 1;  Length 1272;
Best Local Similarity  47.9%;  Pred. No. 0.00049;
Matches 245;  Conservative  0;  Mismatches 265;  Indels  2;  Gaps  2;

Qy   47  cgtttcctcgtgcaatggtgcaccagagcaacgc-cacggcgaggccgcgccgccgc  105
```

```
Db    697  GCCGGCGGTGGCCTCCCCGCGCGAGGCGCCGTGGCAGAGCGACCACCTGTCGCGGGGCGTG  756
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||| |||||||
Qy    473  cctacacaagctatcggccgacgacgtcttcctcaccgcggaaactcaggcgatcgaa  532

Db    757  CCGTACATGCTATCGGCCGACGACGTCTTCCTCACCGCGGCACCGCCGGCACCGACGCCGCCAGGCCAGGC  816
           |||||||||||||||||| | || |||||||| ||||||||| |  || |||  |||||||||
Qy    533  gtcataatcccggtcggttggcccagactgccggcgccaacatactgcttccccgccagc  592

Db    817  GTCATAATCCCGGTGCTGGCCTGGCCGACCACCCGCGCCCGGCCGCCCCAGGCCAGGC  876
           |||||||||||||||||||| ||||||| |||||||||| ||| |||||||
Qy    593  tatccaaattacgagccgacggcattcaacaagtggagtccggcacttcgacctc  652

Db    877  TACCCAAACTACGAGGCCGCGGCCGCCGCGTTCAACAGGCGTGGAGTCCCGCATTTCGACCTC  936
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    653  atccccgacaagggggtgggagatcgacatcgactcgctggaatccatccgccacaagaac  712

Db    937  ATCCCCGACAAGGGGTGGGAGAATCGACATCGACTCGCTGGAATCCATCCGCCGACAAGAAC  996
           |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Qy    713  accaccgatgtcatcataaacccaaacaatccgtcgcagcgtttactcctacgac  772

Db    997  ACCACCGCCATGTCATCATCATAAACCCCAACAACCCGTGCGGCAGCGTTTACTCCCTACGAC  1056
           ||| |||||||||||||||||  |||||| || ||| |||| ||||| || |||||||
Qy    773  catctggccaagtcgccggagtgtggcaaggagctcggaatattggtgatcgtgacgag  832

Db   1057  CATCTGTCCAAGTCGCGGGAGGTGGCGAAAGGCTCGGAATATTGGTGATTGCTGACGAG  1116
           |||||||||||||||||||||||||||||| || |||||||| || | ||||||||
Qy    833  gtttacggcaaactggttctgggcagcgccgccccgtttatcccgatgggcgtctttgggcac  892

Db   1117  GTATACGGCAAGCTGGTTCTGGGCAGCGCCGCCCCGTTCATCCCAATGGAGTGTTTGGGCAC  1176
           ||||| || |||||||| |||| |||||| |||||||| | |||| | ||| || |||||
Qy    893  attgccccggtcttgtccattggatctcgtcaagtcgtggatagtgcctggatgtgcga  952

Db   1177  ATCACCCCTGTCGTGCTGGTCCATAGGGTCTTGTCCAGTCATGGATGATGGCGG  1236
           ||||||| |||| |||||||| ||| |||||||||| ||  |||||||||||||
Qy    953  cttggatgggtggcggtgtacgacccccacagaagatttagagaaactaagatctttacg  1012

Db   1237  CTTGGATGGGTAGCGGGTGTACGACCCCCAGAAAGATCTTACAGAGAAACTAAGAGTCTCACA  1296
           ||||||||||| | ||||||| |||||||| || |||||| || || |||||  ||||||||
Qy   1013  tctattcgattacctaatgtctcaacggaccagcaacttcgttcaggaagctctt  1072

Db   1297  TCAATTACGAATTACCTCCAATGTCTCGACAGACCCAGCAACCTTCATTCAGGCAGCTCTT  1356
           ||||| ||||| ||||||| |||||||| || |||||| |||  ||| |||| || ||||
Qy   1073  cctaaaattcttgagaacacacaaagcagatttcttaagaagattattggtctactaaag  1132

Db   1357  CCTCAGATTCTTGAGAACACAACACAAAGGAAGATTTCTTTAAGGCGATTATTGGTCTGCTAAAG  1416
           |||| |||||||||||||||| ||||||||| ||||||| || | |||| |||||| ||||||
Qy   1133  gaatcatcagagatatgttataggggataataaaagggaaacaaataatatattacgtgtcctcac  1192

Db   1417  GAATCATCAGAGATATGCTACAAACAACAATAAAGGAAACAAAATACAATTACACATGTCCTCAC  1476
           ||||||||||||||| || || ||||||||| |||||| ||| |||| |||||||||| |||||
Qy   1193  aagccagaaggaatcgatgttgtaatggtcaaactaaactaaactttggagggatc  1252

Db   1477  AAGCCAGAAGGAGATCAATGTTGTCATGGTGAAACTGAAACTTACATCTTTTGGAGGAAATA  1536
           ||| ||||| || ||||| || ||  |||| ||||| |||||| ||| |||| || ||||
Qy   1253  catgacgacatagattcttgctcgcaagctcgcaaagggaagaatcagtaatttatgtcca  1312

Db   1537  GACGATGACATAGATTTTTGCTCGCAAGCTCGCAAAAGAAGGAAGAATCAGTAATCTTATGCCCA  1596
           |||||||||| || || |||||||| |||||||||| || || ||| |||| || ||||| |||
Qy   1313  gggagtgttcttggaatggaaaattgggtccgtattactttggctgcgttccatcttct  1372

Db   1597  GGGAGTGTTCTTGGAATGGCAAACTGGGTCCCCATTACTTTGTTGCTTGTTCCATCTTCT  1656
           |||||||||||||||||||| |||| |||||  ||||||||| || ||| |||||||||
Qy   1373  cttcaagatggactcgaaaggtcaaatcattctgtcaaaggaacaagaagaagaattct  1432

Db   1657  CTTCAAGATGTGCTCGGAAGGATCAAATCAATTCATTCTGTCAAAGGAAGAAGAGAAATTCG  1716
           ||||||||||| ||||| ||| ||||||| | |||| || |||||||| || ||| |||||
Qy   1433  ataaatggttgttagttgtacacaccccctagttgtcatctgactgaagctgtaaatcat  1492

Db   1717  AGCGGATCGATTG---------------CTAGTTGTATATCTGACTGACTGAAGCTGTAAAATCAT  1760
           || |||| |||||               |||| || | || |||||||||||||| ||||| |||
Qy   1493  ttctagttatccccattatatattcaataaaacatattgtaatggttcctgtgttagc  1552
```

Query Match   3.8%;   Score 63.8; DB 1;   Length 4524;
Best Local Similarity   50.4%;   Pred. No. 0.00068;
Matches 207; Conservative 0; Mismatches 202; Indels 2; Gaps 2;

```
Qy       47 cgtttcctcgtgcaatgtgtacaccagagcaacgc-cacggcgaggccgccgccgcgc 105
            ||| ||| ||  |  | || || || |  ||  || |||||||||||||||||||||||
Db      965 CGTGTCCGAGTGCCAGATGGACGCCAGCCCGCTGCGCCGCCACGGCCGCCGTGGTGCGC 906

Qy      106 cgccaacggcaagagcaacggcacgcgccccgccgcgcgcgcaagacggcaacggcacgc 165
            || |||  ||  | ||  ||  ||  || ||||||||||||||||||||||||||||||
Db      905 TGCCGCTGCCACCTGCCCCGGGGTGCGGCAGCCCGACGGCCATGTTCATGCGCGTGTGCG 846

Qy      166 gggggcgaagaacagcatccgggcgatacggtacaagatcagcgcgagcgtggaggagag 225
            || ||  ||  |  | | |||  || || ||||| | || || ||| ||| ||| |  |
Db      845 CGCCGGCCCGCCACCAGCCCGGGGTGCGGCAGCCCGACGGCCGCATGTTCATGGCGCCGC 786

Qy      226 ggggcggagaacagcatccgggcgatacggtacaagatcagcgcgagcgtggaggagag 285
            ||  |||  | | |  ||||  ||  || |||||||||||||| |||  || |||||||
Db      785 CGCGGGGTGCGACAGGTGCCCAGGCTGCATGCCCGGGTGCGCGGAGCCCGC 726

Qy      286 cggcgcggccccgtgctgcgcgctggccaccggcaccttccg 345
            ||| || |||||||||||||||||||| | ||||||||| ||
Db      725 CAGGAGCCCGCCCGGGCCCGGGCCCGCC-CGCCCCCGCCTCCCC 667

Qy      346 ctacgcgccgcgtcgcctcccgcgcacgaagcgccgtagcagagcacttgtcaca 405
            | ||| || ||||||||  |  || | |||| | |  ||||||||||| |  ||||
Db      666 CGGGGCCGCCGCGCCGCCCCGCCCGGGCCGCCGCCCCCGGGGCCGCCGGCCCGT 607

Qy      406 ctacgcgccgcgggtggccgcggcgacgtcttcctcaccgcccgcgagcagc 456
            | ||| || || || || |||||||| | |  | |  ||||||||| | | |
Db      606 CGTGGGCGGCCGCCGCCGCCGCCGCCGCCGCCGCCGGGGGCCGCCA 556
```

RESULT 5
V21187

ID   V21187 standard; DNA; 53789 BP.
AC   V21187;
DT   24-JUL-1998 (first entry)
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
OS   Amycolatopsis mediterranei.
FH   Key   Location/Qualifiers
FT   CDS   1825. .15543
FT   /tag= a
FT   /label= ORF_A
FT   /product= "polyketide synthase"
FT   CDS   15550. .30759
FT   /tag= b
FT   /label= ORF_B
FT   /product= "polyketide synthase"
FT   CDS   30895. .36060
FT   /tag= c
FT   /label= ORF_C
FT   /product= "polyketide synthase"
FT   CDS   36259. .41325
FT   /tag= d
FT   /label= ORF_D
FT   /product= "polyketide synthase"
FT   CDS   41373. .51614
FT   /tag= e
FT   /label= ORF_E
FT   /product= "polyketide synthase"
FT   CDS   51713. .5293
FT   /tag= f
FT   /label= ORF_F
FT   /product= "polyketide synthase"
PN   WO9807868-A1.
PD   26-FEB-1998.
PF   18-AUG-1997; E04495.
PR   20-AUG-1996; EP-810551.
PA   (NOVS) NOVARTIS AG.
PI   Engel N, Schupp T, Toupet C;

RESULT 4
V33912/c

ID   V33912 standard; cDNA; 4524 BP.
AC   V33912;
DT   12-FEB-1999 (first entry)
DE   Nucleotide sequence of the SIAX DP2-64 (Oct-T1) gene.
KW   Tumour rejection antigen precursor; TRAP; TRH; leukaemia; screening;
KW   lymphoma; cancer; HLA; human lymphocyte antigen; vaccine;
KW   SIAX DP2-64; Oct-T1; ss.
OS   Homo sapiens.
FH   Key   Location/Qualifiers
FT   CDS   174. .1436
FT   /tag= a
PN   WO9849299-A1.
PD   05-NOV-1998.
PF   22-APR-1998; U07784.
PR   25-APR-1997; US-845998.
PA   (LUDW-) LUDWIG INST CANCER RES.
PI   Boon-Falleur T, Coulie PG, De Smet C, Lucas S, Van Baren N;
DR   WPI; 99-009425/01.
DR   P-PSDB; W68541.
PT   New diagnosis of leukaemia - by detecting genes for tumour antigen
PT   rejection precursors or corresponding proteins
PS   Claim 1: Pages 63-57; 88pp; English.
CC   The present sequence represents the nucleotide sequence of the
CC   SIAX DP2-64 (Oct-T1) gene. The protein is a tumour rejection
CC   antigen precursor (TRAP). The specification describes the treatment of
CC   disorders which characterised by expression of a leukaemia-associated
CC   nucleic acid such as TRH. The products are used for in vivo or in vitro
CC   screening for leukaemia, lymphoma or other cancers by usual
CC   hybridisation/amplification or immunoassay methods. TRAPs, when
CC   processed to antigens or complexed with HLA (human lymphocyte antigen)
CC   molecules, or nucleic acid encoding them, are useful in vaccines for
CC   treating leukaemia.
SQ   Sequence 4524 BP;   1177 A;   1090 C;   1180 G;   1077 T;

```
Db      801 CGTGTCCGAGTGCCAGATGGACGCCAGCCCGCTGCGCCGCCACGGCCGCCGTGGTGCGC 742
            ||| ||| ||  |  | || || || |  ||  ||
Qy      106 cgccaacggcaagagcaacggcacgcgccccgccgcgcgcgcaagacggcaacggcacgc 165
            |||  |  ||  | ||  ||
Db      741 TGCCGCTGCCACCTGCCCCGGGGTGCGGCAGCCCGACGGCCATGTTCATGCGCGTGTGCG 682

Qy      166 gggggcgggcggcggtggagtggaatttcgccgggcaaggacggcatcctgcgacgac 225
            || || || || || ||  ||  | ||  | |  ||  ||||| || |  ||
Db      681 CGCCGCCGCCACCAGCCCGGGGTGCGGCAGCCCGACGGCCATGTTCATGGCGCCGCGC 622

Qy      226 ggggcgaagaacagcatccgggcgatacggtacaagatcagcgcgagcgtggaggagag 285
            |  |||||| || |  || |||  ||  || ||||| | || ||  || |||| |||||
Db      621 CGCGGGGTGCGACAGGTGCCCCAGGCTGCATGCCCGGGTGCGGGTGCGCGGAGCCCGCC 562

Qy      286 cggccgccgtgctgcgcgctggccacggtgacccgtccgtgttcccggccttccg 345
            ||  |  |  |||||||||||||||||  |  | | |  |||   ||| ||| ||
Db      561 CAGGAGCCCGCCCGGGCCCGGGCCCGCC-CGCCCCCGCCTCCCC 503

Qy      346 cacggtcgagcgagacggccgtgctgcgcgctgcgcaccggcaccggccagttcaactg 405
            ||||  ||||| ||||||||  |||||||||||||||||||
Db      502 CGGGGCCGCCGCGCCGCCCCGCCCGGGCCCGCCCGGGGGCCGT 443

Qy      406 ctacgcgccgcgtcgcctcccgcgcacgaagcgccgtagcagagcacttgtcaca 465
            |  || ||  ||  | | |  | ||  | |||| | |  |||||| |  ||||||
Db      442 CGTGGGCGGCCGCCGCCGCCGCCGCCGGGCCGCCGCCCCGGGGCCGCCA 383

Qy      466 gggcgtgccctacaagctatcgccgacgacgtcttcctcaccgcccgcgggaactcaggc 525
            || ||  || | ||  || ||||||| | | ||  | | ||||||  ||  ||||| |||
Db      382 TGAGCGCGAGCGACGGCGAGGAGATGTGTCCAGCAGATCGCCGGGTTCGAGCGCCTTGGT 323

Qy      526 gatcgaagtcatatatccgggtgctggccccaga 557
            |||||||| |  | |||| ||| ||  || |||||
Db      322 GGTGGTGGTGGTGGTGCTGGTGGTGGTGCCCCAGA 291
```

```
DR   WPI; 98-169172/15.
DR   P-PSDB; W52845-W52850.
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
PS   Claim 4; Page 53-102; 205pp; English.
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
SQ   Sequence  53789 BP;    6707 A;   19183 C;    20504 G;    7395 T;

Query Match             3.7%;   Score 62;  DB 1;   Length 53789;
Best Local Similarity   48.8%;    Pred. NO. 0.0024;
Matches  167;  Conservative   0;  Mismatches 175;  Indels   0;  Gaps   0;

Qy      58 ggcaatggtacaccagagacaacggccacggcggaggccgccgccgccgccaacggcaa 117
           ||   ||||| ||||    || ||||||||| || |||| |||||||||||| |||||
Db   35607 GGAAGCGGCTGCTGCGTCAGACCTCGTCCGCGTCCAGTTCGCGCTCTGCTCGGGCACGCCGA 35666

Qy     118 gagcaacaggcacgcgccgcgcccgaacggcaagaagcacaggcacggcggcggcggc 177
           ||    |||||| || ||||| ||||| ||| ||| || ||||||||| ||| |||||
Db   35667 CAGCTCCGGCGTCCCGCGCCGACGCGGCGTTCAAGGACGCCGGGTTCGACTCGCTGACGTC 35726

Qy     178 ggtggagtggaatttcgcccgggcaaggacggcatcctggcgacgacgggggcgaagaa 237
           |||||||         ||| ||| ||||||||   || |||||||| |||| |||||||
Db   35727 GGTGGAGCTGCGCAACCGGCTGCCCGAGACGACCCGGCCCTGAAACTGCCCGGCGACGCTGGT 35786

Qy     238 cagcatccgggcgatacggtacaagatcagcgcgagcgtggaggagacggggccgcggcc 297
           || || |||    || ||| ||||  |||||| ||||| |||| ||||| ||||| |||||
Db   35787 CTTCGACCATCCGAACCCGCTGGCACTGCCCGCACTGGCCCGGCGGAACTCGCCGTCGA 35846

Qy     298 cgtgctgccgctggccacgtgaccgtccgtgttcccgacgcctccgacggccgtcga 357
           || ||||| ||||| ||| ||| |||||  || ||| || |||| ||||||||||
Db   35847 CGAGGCATCCCCGGCCGATCGGTGCTGGCCGGCTCGCCGGCCGTGGAGGCGGCCATCGC 35906

Qy     358 ggccgaagacgccgtgccgcgccggctgcgcaccggccagtt 399
           || || || ||||| |||| | ||| |||||| ||||| |||
Db   35907 GGCCGCCGGCGGCCGCCCCCGGACACCGGCGCCCCGGATCACCGGCTCACCGGCGCGGCT 35948

RESULT      6
ID   V62134/c
AC   V62134;
DT   07-DEC-1998  (first entry)
DE   HSV-2 strain SB5 Contig ID 104 DNA sequence.
KW   HSV-2 strain SB5; immunological response induction; therapy;
KW   antiviral identification; viral protein inhibitor; ss.
OS   Herpes simplex virus type 2.
FH   Key          Location/Qualifiers
FT   CDS          complement (256..3021)
FT                /*tag= a
FT                /product= "ORF#1 protein"
FT                /note= "encoded protein shown in W72033"
FT   CDS          complement (2895..3908)
FT                /*tag= b
FT                /product= "ORF#2 protein"
FT                /note= "encoded protein shown in W72034"
```

```
FT   CDS          complement (3637..6093)
FT                /*tag= c
FT                /product= "ORF#3 protein"
FT                /note= "encoded protein shown in W72035"
FT   CDS          complement (6263..6745)
FT                /*tag= d
FT                /product= "ORF#4 protein"
FT                /note= "encoded protein shown in W72036"
FT                6775..8046
FT                /*tag= e
FT                /product= "ORF#5 protein"
FT                /note= "encoded protein shown in W72037"
FT   CDS          complement (8386..9477)
FT                /*tag= f
FT                /product= "ORF#6 protein"
FT                /note= "encoded protein shown in W72038"
FT   CDS          complement (9721..13911)
FT                /*tag= g
FT                /product= "ORF#7 protein"
FT                /note= "encoded protein shown in W72039"
FT   CDS          complement (14158..14826)
FT                /*tag= h
FT                /product= "ORF#8 protein"
FT                /note= "encoded protein shown in W72040"
FT                15162..17075
FT                /*tag= i
FT                /product= "ORF#9 protein"
FT                /note= "encoded protein shown in W72041"
FT   CDS          complement (17295..19110)
FT                /*tag= j
FT                /product= "ORF#10 protein"
FT                /note= "encoded protein shown in W72042"
FT   CDS          complement (19111..20064)
FT                /*tag= k
FT                /product= "ORF#11 protein"
FT                /note= "encoded protein shown in W72043"
FT   CDS          complement (20149..21279)
FT                /*tag= l
FT                /product= "ORF#12 protein"
FT                /note= "encoded protein shown in W72044"
FT                20962..22026
FT                /*tag= m
FT                /product= "ORF#13 protein"
FT                /note= "encoded protein shown in W72045"
FT                22307..23887
FT                /*tag= n
FT                /product= "ORF#14 protein"
FT                /note= "encoded protein shown in W72046"
FT                23778..24065
FT                /*tag= o
FT                /product= "ORF#15 protein"
FT                /note= "encoded protein shown in W72047"
FT   CDS          complement (24286..25632)
FT                /*tag= p
FT                /product= "ORF#16 protein"
FT                /note= "encoded protein shown in W72048"
FT                25457..26215
FT                /*tag= q
FT                /product= "ORF#17 protein"
FT                /note= "encoded protein shown in W72049"
PN   WO9820016-A1.
PD   14-MAY-1998.
PF   31-OCT-1997; U20016.
PR   09-JUN-1997; US-049018.
PR   04-NOV-1996; US-030279.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
PI   Esser KM, Leary JJ;
DR   WPI; 98-286847/25.
DR   P-PSDB; W72033, W72034, W72035, W72036, W72037, W72038, W72039, W72040,
DR   W72041, W72042, W72043, W72044, W72045, W72046, W72047, W72048, W72049.
PT   Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT   and treatment of infection or inducing immunological response in
```

PT mammal
PS Claim 1; Page 226-237; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC sequence of the invention. This sequence was isolated from HSV-2 strain
CC SB5 (deposited as ATCC VR-2546), is designated Contig ID 104, and encodes
CC 17 HSV-2 proteins. The proteins can be used for the treatment or
CC prevention of disease, to induce an immunological response in a mammal or
CC to identify inhibitors, activators or novel antivirals. Antagonists of
CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC or a vector containing it can also be used to induce an immunological
CC response in a mammal.
SQ Sequence 26338 BP; 4257 A; 8879 C; 9241 G; 3959 T;

Query Match   3.6%; Score 59; DB 1; Length 26338;
Best Local Similarity 47.4%; Pred. No. 0.008;
Matches 176; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

Qy   57 tggcaatggtacaccagagcaacggccacggcagggccgcgccgaggccgccgccaacggca 116
         ||| |||| || |||||||||| | |||| || |||| ||||||||| |||
Db 2270 TGGGCAAGGTCATCGGCCGCACGCGACGCCCGCCGAGGAACTTGAGGCCCGTCGCCG 2211

Qy   117 agagcaacgggcacgcgccgccgcgcgcgcaacggcaagagcaacgggcacgcggcggcgg 176
         | ||| ||||| ||||||| |||||| | ||| || |||| ||||||| ||| ||||||
Db 2210 ACCTGACGGGGAAGATGGGCGCCCCAGCCGCGCCCAGCGGGCAGCAGCGAGCGCGTGGGCCGCCG 2151

Qy   177 cggtggagtggaatttcgccgcggggcaaggacggcatcctgcgacgacgggggcgaaga 236
         || ||| |||||| | || | | |||||||||| ||| | ||| ||| |||||| ||||
Db 2150 GCGTCGAGGCGGCGGCGCTGGACCCGGCCGTCGAGAACCGCGCCGAGTTCGACGTGGTCGAGCTGC 2091

Qy   237 acagcatccgggcgatacggtacaagatcacgcgagcgtggaggagagccggccgccggc 296
         ||| |||||||| ||| | || | || ||| |||| | |||| | ||| |||
Db 2090 GCCGCCTGCCTGCAGCGGCTGCGGGCACGCACGCTACAACCCCGGACTTCCCGAAGCGCG 2031

Qy   297 ccgtgctgccctggccccacggtgacccgtcgtgttcccgggccttccgcacggcgctcg 356
         ||| ||||| |||||| |||||| ||| | ||| ||| ||||| || | |||| || |
Db 2030 CCGAGCAGGCGCTCGCGGCCAACGCGCCAGGCGGTCACCCTCGCCCCTGGATACGGCCTTCG 1971

Qy   357 aggccgaagacgggcgtcgccgcgcgctgcgcaccggcagttcaactgctacggccgccg 416
         | | | ||| ||| | | || | | | | || || | ||||| || ||| || |||
Db 1970 CGTTTAACCCCTACACCCCGAGAACCAGCGCCACCCCATGCTCCCCCCGCGCCGCCA 1911

Qy   417 ggctcggcctc 427
         ||||||||||
Db 1910 TCCACCGCCTC 1900

RESULT 7
V62154/c
ID V62154 standard; DNA; 21034 BP.
AC V62154;
DT 18-DEC-1998 (first entry)
DE HSV-2 strain SB5 Contig ID 10 DNA sequence.
KW HSV-2 strain SB5; immunological response induction; therapy;
KW antiviral identification; viral protein inhibitor; ss.
OS Herpes simplex virus type 2.
FH Key   Location/Qualifiers
FT CDS   complement (1..5672)
FT   /*tag= a
FT   /transl_except= (pos: 1..2, aa: Thr-Xaa-Xaa-Xaa-Xaa)
FT   /product= "ORF#1 protein"
FT   /note= "encoded protein shown in W72095, Xaa= unspecified
FT   amino acid"
FT CDS   complement (5962..9306)
FT   /*tag= b
FT   /product= "ORF#2 protein"
FT   /note= "encoded protein shown in W72096"
FT   9246..11195
FT   /*tag= c
FT CDS   /product= "ORF#3 protein"
FT   /note= "encoded protein shown in W72097"
FT   11657..15199
FT   /*tag= d

FT   /product= "ORF#4 protein"
FT   /note= "encoded protein shown in W72098"
FT   15252..16265
FT CDS   /*tag= e
FT   /product= "ORF#5 protein"
FT   /note= "encoded protein shown in W72099"
FT   complement (16548..17621)
FT   /*tag= f
FT   /product= "ORF#6 protein"
FT CDS   /note= "encoded protein shown in W72100"
FT   18521..19936
FT   /*tag= g
FT   /product= "ORF#7 protein"
FT CDS   /note= "encoded protein shown in W72101"
FT   20015..21034
FT   /*tag= h
FT   /transl_except= (pos: 21033..21034, aa: Gly-Ser)
FT CDS   /product= "ORF#8 protein"
FT   /note= "encoded protein shown in W72102"
PN WO9820016-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20016.
PR 09-JUN-1997; US-049018.
PR 04-NOV-1996; US-030279.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Chan JY, Dabrowski-Amaral CE, Delvecchio AM, Dillon SB,
PI Esser KM, Leary JJ;
DR WPI; 98-286847/25.
DR P-PSDB; W72095, W72096, W72097, W72098, W72099, W72100, W72101, W72102.
PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT and treatment of infection or inducing immunological response in
PT mammal
PS Claim 1; Page 350-359; 748pp; English.
CC This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC sequence of the invention. This sequence was isolated from HSV-2 strain
CC SB5 (deposited as ATCC VR-2546), is designated Contig ID 10, and encodes
CC 8 HSV-2 proteins. The proteins can be used for the treatment or
CC prevention of disease, to induce an immunological response in a mammal or
CC to identify inhibitors, activators or novel antivirals. Antagonists of
CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC or a vector containing it can also be used to induce an immunological
CC response in a mammal.
SQ Sequence 21034 BP; 3154 A; 7419 C; 7234 G; 3227 T;

Query Match   3.6%; Score 59; DB 1; Length 21034;
Best Local Similarity 47.4%; Pred. No. 0.0076;
Matches 176; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

Qy   57 tgtcaatggtacaccagagcaacggccacggcagggccgcgccgaggccgccgccaacggca 116
         ||| |||| || |||||||||| | |||| || |||| ||||||||| |||
Db 1780 TGGGCAAGGTCATCGGCCGCACGCGACGCCCGCCGAGGAACTTGAGGCCCGTCGCCG 1721

Qy   117 agagcaacgggcacgcgccgccgcgcgcgcaacggcaagagcaacgggcacgcggcggcgg 176
         | ||| ||||| ||||||| |||||| | ||| || |||| ||||||| ||| ||||||
Db 1720 ACCTGACGGGGAAGATGGGCGCCCCAGCCGCGCCCAGCGGGCAGCAGCGAGCGCGTGGGCCGCCG 1661

Qy   177 cggtggagtggaatttcgccgcggggcaaggacggcatcctgcgacgacgggggcgaaga 236
         || ||| |||||| | || | | |||||||||| ||| | ||| ||| |||||| ||||
Db 1660 GCGTCGAGGCGGCGGCGCTGGACCCGGCCGTCGAGAACCGCGCCGAGTTCGACGTGGTCGAGCTGC 1601

Qy   237 acagcatccgggcgatacggtacaagatcacgcgagcgtggaggagagccggccgccggc 296
         ||| |||||||| ||| | || | || ||| |||| | |||| | ||| |||
Db 1600 GCCGCCTGCCTGCAGCGGCTGCGGGCACGCACGCTACAACCCCGGACTTCCCGAAGCGCG 1541

Qy   297 ccgtgctgccctggccccacggtgacccgtcgtgttcccgggccttccgcacggcgctcg 356
         ||| ||||| |||||| |||||| ||| | ||| ||| ||||| || | |||| || |
Db 1540 CCGAGCAGGCGCTCGCGGCCAACGCGCCAGGCGGTCACCCTCGCCCCTGGATACGGCCTTCG 1481

Qy   357 aggccgaagacgggcgtcgccgcgcgctgcgcaccggcagttcaactgctacggccgccg 416
         | | | ||| ||| | | || | | | | || || | ||||| || ||| || |||
Db 1480 CGTTTAACCCCTACACCCCGAGAACCAGCGCCACCCCATGCTCCCCCCGCGCCGCCA 1421

Fri Oct 1 16:47:10 1999

us-09-026-400-1.rng

```
Qy    417 ggtcggcctc 427
          | | ||||
Db   1420 TCCACCGCCTC 1410

RESULT     8
V62176
ID   V62176 standard: DNA; 117213 BP.
AC   V62176;
DT   13-JAN-1999 (first entry)
DE   HSV-2 strain SB5 Contig ID 15 DNA sequence.
KW   HSV-2 strain SB5; immunological response induction; therapy;
KW   antiviral identification; viral protein inhibitor; ss.
OS   Herpes simplex virus type 2.
FH   Key             Location/Qualifiers
FT   CDS             755..1297
FT                   /*tag= a
FT                   /product= "ORF#1 protein"
FT                   /note= "encoded protein shown in W72170"
FT   CDS             1170..2174
FT                   /*tag= b
FT                   /product= "ORF#2 protein"
FT                   /note= "encoded protein shown in W72171"
FT   CDS             2229..2930
FT                   /*tag= c
FT                   /product= "ORF#3 protein"
FT                   /note= "encoded protein shown in W72172"
FT   CDS             complement (3130..3735)
FT                   /*tag= d
FT                   /product= "ORF#4 protein"
FT                   /note= "encoded protein shown in W72173"
FT   CDS             complement (3802..6447)
FT                   /*tag= e
FT                   /product= "ORF#5 protein"
FT                   /note= "encoded protein shown in W72174"
FT   CDS             6017..8482
FT                   /*tag= f
FT                   /product= "ORF#6g protein"
FT                   /note= "encoded protein shown in W72250"
FT   CDS             6026..8482
FT                   /*tag= g
FT                   /product= "ORF#6f protein"
FT                   /note= "encoded protein shown in W72249"
FT   CDS             6065..8482
FT                   /*tag= h
FT                   /product= "ORF#6e protein"
FT                   /note= "encoded protein shown in W72248"
FT   CDS             6167..8482
FT                   /*tag= i
FT                   /product= "ORF#6d protein"
FT                   /note= "encoded protein shown in W72247"
FT   CDS             6296..8482
FT                   /*tag= j
FT                   /product= "ORF#6c protein"
FT                   /note= "encoded protein shown in W72246"
FT   CDS             6326..8482
FT                   /*tag= k
FT                   /product= "ORF#6b protein"
FT                   /note= "encoded protein shown in W72245"
FT   CDS             6446..8482
FT                   /*tag= l
FT                   /product= "ORF#6a protein"
FT                   /transl_except= (pos: 7400..7402, aa: Ala-Ala)
FT                   /transl_except= (pos: 7481..7486, aa: Ile)
FT                   /note= "encoded protein shown in W72244"
FT   CDS             8457..9347
FT                   /*tag= m
FT                   /product= "ORF#7 protein"
FT                   /note= "encoded protein shown in W72175"
FT   CDS             complement (9604..11855)
FT                   /*tag= n
FT                   /transl_except= (pos: 11635..11636, aa: Ala)
FT                   /product= "ORF#8 Protein"

FT                   /note= "encoded protein shown in W72176"
FT                   complement (11905..14508)
FT                   /*tag= o
FT                   /product= "ORF#9b protein"
FT                   /note= "encoded protein shown in W72222"
FT   CDS             complement (11905..14520)
FT                   /*tag= p
FT                   /product= "ORF#9a protein"
FT                   /note= "encoded protein shown in W72223"
FT                   14399..15802
FT                   /*tag= q
FT                   /product= "ORF#10 protein"
FT                   /note= "encoded protein shown in W72177"
FT   CDS             complement (15996..16286)
FT                   /*tag= r
FT                   /product= "ORF#11 protein"
FT                   /note= "encoded protein shown in W72178"
FT   CDS             complement (16202..18064)
FT                   /*tag= s
FT                   /product= "ORF#12 protein"
FT                   /note= "encoded protein shown in W72179"
FT                   complement (18105..19661)
FT                   /*tag= t
FT                   /product= "ORF#13 protein"
FT                   /note= "encoded protein shown in W72180"
FT   CDS             complement (19415..20074)
FT                   /*tag= u
FT                   /product= "ORF#14 protein"
FT                   /note= "encoded protein shown in W72181"
FT                   20155..21453
FT                   /*tag= v
FT                   /product= "ORF#15 protein"
FT                   /note= "encoded protein shown in W72182"
FT   CDS             complement (21326..22291)
FT                   /*tag= w
FT                   /product= "ORF#16 protein"
FT                   /note= "encoded protein shown in W72183"
FT   CDS             complement (22546..24654)
FT                   /*tag= x
FT                   /product= "ORF#17 protein"
FT                   /note= "encoded protein shown in W72184"
FT                   24684..25955
FT                   /*tag= y
FT                   /product= "ORF#18 protein"
FT                   /note= "encoded protein shown in W72185"
FT   CDS             complement (26295..27251)
FT                   /*tag= z
FT                   /product= "ORF#19 protein"
FT                   /note= "encoded protein shown in W72186"
FT   CDS             complement (27630..31754)
FT                   /*tag= aa
FT                   /product= "ORF#20b protein"
FT                   /note= "encoded protein shown in W72225"
FT                   complement (27630..31784)
FT                   /*tag= ab
FT                   /product= "ORF#20a protein"
FT                   /note= "encoded protein shown in W72224"
FT   CDS             complement (32067..32735)
FT                   /*tag= ac
FT                   /product= "ORF#21 protein"
FT                   /note= "encoded protein shown in W72187"
FT                   33140..34984
FT                   /*tag= ad
FT                   /product= "ORF#22a protein"
FT                   /note= "encoded protein shown in W72226"
FT                   33386..34984
FT                   /*tag= ae
FT                   /product= "ORF#22b protein"
FT                   /note= "encoded protein shown in W72227"
FT   CDS             complement (35205..37721)
FT                   /*tag= af
FT                   /product= "ORF#23 protein"
FT                   /note= "encoded protein shown in W72188"
```

```
FT   CDS             complement (38058..39188)
FT                   /*tag= ag
FT                   /product= "ORF#24 protein"
FT                   /note= "encoded protein shown in W72189"
FT   CDS             39090..39935
FT                   /*tag= ah
FT                   /product= "ORF#25 protein"
FT                   /note= "encoded protein shown in W72190"
FT   CDS             40216..41973
FT                   /*tag= ai
FT                   /product= "ORF#26 protein"
FT                   /note= "encoded protein shown in W72191"
FT   CDS             42206..44178
FT                   /*tag= aj
FT                   /transl_except= (pos: 44063..44064, aa: Lys)
FT                   /product= "ORF#27 protein"
FT                   /note= "encoded protein shown in W72192"
FT   CDS             complement (44853..47297)
FT                   /*tag= ak
FT                   /product= "ORF#28 protein"
FT                   /note= "encoded protein shown in W72193"
FT   CDS             47122..47338
FT                   /*tag= al
FT                   /product= "ORF#29 protein"
FT                   /note= "encoded protein shown in W72194"
FT   CDS             complement (47305..49662)
FT                   /*tag= am
FT                   /product= "ORF#30 protein"
FT                   /note= "encoded protein shown in W72195"
FT   CDS             complement (50035..51666)
FT                   /*tag= an
FT                   /product= "ORF#31 protein"
FT                   /note= "encoded protein shown in W72196"
FT   CDS             complement (51701..53575)
FT                   /*tag= ao
FT                   /product= "ORF#32 protein"
FT                   /note= "encoded protein shown in W72197"
FT   CDS             54393..58115
FT                   /*tag= ap
FT                   /product= "ORF#33 protein"
FT                   /note= "encoded protein shown in W72198"
FT   CDS             complement (58060..58977)
FT                   /*tag= aq
FT                   /product= "ORF#34 protein"
FT                   /note= "encoded protein shown in W72199"
FT   CDS             complement (58970..60760)
FT                   /*tag= ar
FT                   /product= "ORF#35 protein"
FT                   /note= "encoded protein shown in W72200"
FT   CDS             60759..61151
FT                   /*tag= as
FT                   /product= "ORF#36 protein"
FT                   /note= "encoded protein shown in W72201"
FT   CDS             61241..62071
FT                   /*tag= at
FT                   /product= "ORF#37 protein"
FT                   /note= "encoded protein shown in W72202"
FT   CDS             62183..62521
FT                   /*tag= au
FT                   /product= "ORF#38 protein"
FT                   /note= "encoded protein shown in W72203"

Query Match             3.6%;  Score 59;  DB 1;  Length 117213;
Best Local Similarity  47.4%;  Pred. No. 0.011;
Matches  176;  Conservative   0;  Mismatches  195;  Indels   0;  Gaps   0;

Qy      57 tgcaatgtacaccagagcaacggcccacgccgaggccgccgccgccgccaacggca 116
           | |  |  | | || ||||  |||| | ||| |||| ||||||||||||||
Db   68173 TGGGCAAGGTCATCGGCGCCACGCGACCGCGCCGCCGACGAACTTGAGCGCCGCTCGCCG 68114

Qy     117 agagcaacggcacgcgccgcgcgaacggcaagagagcaacggcacgcggcgggcgg 176
           | ||  |  | ||||  | | |  |  | ||  | || | |  | |||||| |||
Db   68113 ACCTGACGGCGAAGATGGGCGCCCCAGCGCGCGGGCAGCAGCGAGCGCTGGGCGGCCG 68054
```

```
Qy     177 cggtggatggaatttcgcccgggcaaggacggcatcctggcgacgacgggggcgaaga 236
           |||  |||  ||    |||| ||| |||||||  ||| | | ||||||| | ||| ||
Db   68053 GCGTCGAGCGGCGGCGTGACCGGCTGGAACCGGCCGAGTTCGACGTGGTCGAGCTGC 67994

Qy     237 acagcatccggcggcgatacgtacaagatcagcggcgagcgtgaggagagcggccggggc 296
           |||||||| ||| || ||||||| |||| ||| || ||| ||||||||||  |||||||
Db   67993 GCCGCCTGCAGGCGGCTGCGGGCACGCGACGCTACAACCCCGGACTTCCGCAAGCGCCG 67934

Qy     297 ccgtgcgtgccgctggccacggtgacccgtccgtgttccggccttccgcacggccgtcg 356
           ||||| ||||  ||||||||||| |  || ||  ||  | |||| ||||  |||| || |
Db   67933 CCGAGCAGGCGGCTTCGGCGCCAACGCGCGGAGGCGGTCACCTGCGCCCTGGATACGGCCTTCG 67874

Qy     357 agccgagaagaccgccgtgcgcgcgggctgcgcgacggccgagttcaactgctacgccgcg 416
           || | | |  |||||| ||||| | | | |  ||| ||||| ||| || ||||| |||
Db   67873 CGTTAACCCTACACCCCGAGAACCAGCGCCACCCCCGCTCCCCCCGCTCGCCGCCCGCCA 67814

Qy     417 ggctcggcctc 427
           |  | | |||
Db   67813 TCCACCGCCTC 67803
```

```
RESULT   9

X09011/c
ID   X09011 standard; cDNA; 1266 BP.
AC   X09011;
DT   14-JUN-1999  (first entry)
DE   Brn-3a polynucleotide.
KW   Brn-3a; Bcl-2; neurons; neuronal cells; apoptosis; cell death; CNS;
KW   PNS; central nervous system; parasympathetic nervous system;
KW   development; injury; neurotrophic factor; nerve growth factor; NGF;
KW   ciliary neurotrophic factor; CNTF; brain-derived neurotrophic factor;
KW   BNTF; neurotrophin; NT-3; NT-4; NT-5; neurodegenerative disease;
KW   familial dysautonomia; infantile muscular dystrophy;
KW   Parkinson's disease; Alzheimer's disease; ss.
OS   Mus musculus.
FH   Key            Location/Qualifiers
FH
FT   CDS            1..1266
FT                  /*tag= a
FT                  /product= Brn_3a_polypeptide
PN   WO9905272-A1.
PD   04-FEB-1999.
PF   27-JUL-1998; G02228.
PR   10-DEC-1997; US-988476.
PR   25-JUL-1997; GB-015823.
PA   (UNLO ) UNIV COLLEGE LONDON.
PI   Latchman DS, Smith MD;
DR   WPI; 99-142928/12.
DR   P-PSDB; W96263.
PT   New polypeptide comprising transcription factor Brn-3a, or its
PT   derivative - useful for treating nervous system diseases, preventing
PT   cellular apoptosis and increasing nerve regeneration following
PT   neuronal damage
PS   Disclosure; Page 63-64; 68pp; English.
CC   Over expression of transcription factor Brn-3a can protect neuronal
CC   cells from apoptosis. Brn-3a also specifically activates expression
CC   of the Bcl-2 gene in neuronal cells and this activation is mediated
CC   via a Brn-3a response element in the 5' regulatory region of the
CC   Bcl-2 gene. Both the anti-apoptopic effect of Brn-3a and its ability
CC   to activate expression of Bcl-2 are mediated by the N-terminal
CC   domain of Brn-3a. Members of the Bcl-2 family perform critical roles
CC   in the regulation of selective apoptosis during development of the
CC   nervous system. The stimulation of Bcl-2 expression by Brn-3a in a
CC   neuron specific manner and consequent protection of neuronal cells
CC   from apoptosis suggests that Brn-3a may co-ordinate some aspects of
CC   neuronal reorganisation during development or following injury. The
CC   elevation of Brn-3a expression by either pharmacological means
CC   (compositions comprising one other therapeutic polypeptide e.g.
CC   neurotrophic factors, nerve growth factor (NGF), ciliary neurotrophic
CC   factor (CNTF), brain-derived neurotrophic factor (BNTF), and
CC   neurotrophins NT-3 and NT-4/5) or gene therapy may represent a
CC   method for treating human diseases associated with excessive
CC   neuronal cell death and/or lack of nerve regeneration, especially
```

```
CC   neurodegenerative diseases such as familial dysautonomia and
CC   infantile muscular dystrophy, and Parkinson's and Alzheimer's
CC   disease.
SQ   Sequence  1266 BP;  201 A;  455 C;  458 G;  152 T;

Query Match    3.5%;    Score 58;  DB 1; Length 1266;
Best Local Similarity 49.4%;  Pred. No. 0.0069;
Matches  177; Conservative  0; Mismatches 180; Indels  1; Gaps  1;

Qy    93 ccgccgccgcgccaacgcgcaagagcaacggcgccgccgccaacggcaagagcaacggcaaga 152
            ||||||||||| ||  |||||||||||  ||||| ||||||
Db   754 CCACCGCCGCGCGACGGCGCGCCGAGCGGCGCGCCCGCCGCCGCCCGCCGCCGCTGCCG 695

Qy   153 gcaacgggcacgcgcggcggcgggtggagtgaattcgccgggcgggcaaggacggca 212
            ||||  |||||  |||||||||||||  |||| ||||||
Db   694 CGCCGCCGCGCTGGTGCCCGCCGCCACGAGCCGCGGATGCGGCACGAGCCCGGACGGCA 635

Qy   213 tcctggcgacgacgggggcgaagaacagcatccgggcgatacacggtacaagatcagcgcga 272
            |||| ||||||||   |||||||||||||||||
Db   634 TGTTCATGGCCCGCCGCGCGCGGGGTGCGACAGGTGCCCAGGCCGTGCATGTGCGGGT 575

Qy   273 gcgtggaggagagcgggcgcgcccgtgctgcgccgctgctgcctgcggcgctcgtgt 332
            |||  ||||| |||||
Db   574 GCGGATGCGGCCGAGCCCCGGGGCGCCCCAAGAGCCCCCGGGCGCCCCGGGGGGC-CGCCGCCCCCGGG 516

Qy   333 tcccggccttccgacggccgtcgaggccgaagacgccgtcgcggcgctgcggcgctgcgccaccg 392
            ||||| |||| |||  ||||||||||  |||| ||  |||||||||
Db   515 CCGGCCGCCGCGCGCCGGCCGCGCCACCCCGGGGCGGCCGCGCGCCCACCGCCCCCGGT 456

Qy   393 gccagttcaactgctacgcgccgccggtcggcctcccccgcgcacgaagcgccgtagc 450
            ||||||  |||| |||  ||||  ||||||||||||
Db   455 CGCCTCCCCGCCCCCGGGGGCCGTCGTGGGGCGGCCGCCTCCCCGCCGCCCCTGCGC 398
```

```
RESULT 10
N92408
ID   N92408 standard; DNA; 2061 BP.
AC   N92408;
DT   11-MAR-1992 (first entry)
DE   Sequence encoding pseudorabies virus (PRV) glycoprotein gH.
KW   Herpes virus; vaccine; surface polypeptide; Bulbar paralysis;
KW   Aujeszky's disease; mad itch; tracheitis; rhinitis; ss.
OS   Pseudorabies virus.
FH   Key             Location/Qualifiers
FT   cds             1..2061
FT                   /*tag=  a
PN   WO8910965-A.
PD   16-NOV-1989.
PF   18-APR-1989; 901563.
PR   03-MAY-1988; US-189907.
PA   (UPJO ) UPJOHN CO.
PI   Petrovskis EA, Post LE;
DR   WPI; 89-356494/48.
DR   P-PSDB: P93711.
PT   Recombinant DNA - comprises sequence encoding glycoprotein gH of
PT   pseudorabies virus or infectious bovine rhinotracheitis virus
PS   Claim 2; Page 24 and Chart A, Page 15-17; 31pp; English.
CC   The recombinant DNA of the invention (see N92408 and N93701) can be
CC   used to diagnose PRV and IBRV infections. The encoded proteins
CC   (see P93711 and P93712) or their fragments or derivs., are useful as
CC   virus-free vaccine components for protecting animals against PRV or
CC   IBRV infections. Host cells transformed with N92408 or N93701
CC   operably linked to an expression control sequence are also claimed.
CC   The host cell is from a bacteria, fungus plant, or animal cell. Pref.
CC   the host is an insect cell infected with a baculovirus.
SQ   Sequence  2061 BP;  199 A;  846 C;  727 G;  289 T;

Query Match    3.4%;    Score 56.4;  DB 1; Length 2061;
Best Local Similarity 48.7%;  Pred. No. 0.015;
Matches  153; Conservative  0; Mismatches 161; Indels  0; Gaps  0;
```

```
Qy    75 gcaacggccaacgccgaggccgccgccgccgccaacggcaagagacaacggcacgccg 134
           || ||||| |||  ||||||||||||||||||||||||||| |||||||||
Db  1166 GCCCCGGCGACGGCGGCGCGGAGAACCTGCTCCGCACCACGGCCATGTCACCGGAGCGCG 1225

Qy   135 ccgcccgaacgcaagagcaacggcggcgggggcggcgggtggagtgaaattcg 194
           |||||||  ||||||||| |||||||||
Db  1226 CCGGCGGCGGCGGCCGAGCTCACGCGGCCGGCCTCCGCCGGCGGCGCGTGGAACGAGC 1285

Qy   195 ccgggcaaggacggcatcctggcgacgacgggcgaagacagcatccgggcgatac 254
           |||||  || ||| ||
Db  1286 CCTTCAGCCTGCTGACGTCCTCTCGCCGCCGGTCTCGCCTGCCCGCGACCTCGGCG 1345

Qy   255 ggtacaagatcagcggcgagcgtggaggagagcgggcgcgcgccgtgctgcgctgcccc 314
           ||||
Db  1346 GGGACGCGCACCCTGGCCAACCTGGCGCCGCCGCGCGGCGTCCGCCTGCGCCTGCGGGG 1405

Qy   315 acggtgacccgtccgttcccggccttccgcacggccgtcgaggccgaagacgccgtcg 374
           || |||
Db  1406 CCCCGGCGGCGGCGGGCGGACGACGAGGAGGGCGGAGGAGGAGGACCCCGTCG 1465

Qy   375 ccgcggcgctgcgc 388
           ||| |||
Db  1466 CGGCGGCGCGCCCC 1479
```

```
RESULT 11
X53491/c
ID   X53491 standard; DNA; 114955 BP.
AC   X53491;
DT   05-JUL-1999 (first entry)
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
OS   Synthetic.
PN   WO9913886-A1.
PD   25-MAR-1999.
PF   17-SEP-1998; U19419.
PR   09-JUN-1998; US-093972.
PR   17-SEP-1997; US-059160.
PA   (UYEC-) UNIV EAST CAROLINA.
PI   Nyce JW;
DR   WPI; 99-229400/19.
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
PS   Disclosure; Page 37; 120pp; English.
CC   The specification describes antisense oligonucleotides (X52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences X55272-74. These multiple target
CC   oligonucleotides (specifically X55180-271) can be used for the antisense
CC   treatment of diseases and conditions. Typical diseases and conditions
CC   are those associated with impaired respiration and inflammation,
CC   including lung diseases, pulmonary vasoconstriction, inflammation,
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC   hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC   metastases, as well as all types of cancers which may metastasize or have
CC   metastasized to the lungs, including breast and prostate cancer.
```

SQ   Sequence 114955 BP;   6071 A;   29417 C;   36712 G;   21328 T;

Query Match   3.4%;   Score 56.2; DB 1; Length 114955;
Best Local Similarity 34.0%;   Pred. No. 0.037;
Matches 131; Conservative 40; Mismatches 214; Indels 0; Gaps 0;

```
Qy      132 cgcgccgggccaagagagcaagaggcacggcacgcggcgggcgggcggcggtggagtggagaatt 191
            |:||| |:|| : || ||        ||: || |:||:|:: |||:
Db   104788 CVGCGGNNHNNNSCGGCCVVGCCVGCGGNNHNNNSCGGCCVGCCVGCGGNNHNNNSCG 104729

Qy      192 tcgccccgggccaagacggcatcctggcgacgacggggcggcgaagaacagcatccgggcga 251
            |:|| : || | ||  :|:||:  ||
Db   104728 CGCCCCVGCCVGCGGNNHNNNSGGGCGGCCVGCCVGCGGNNHNNNSGGCGG 104669

Qy      252 tacggtacaagatcgcgagcgggaggagagcggcccgtgccgctgg 311
            : |||| ||   :|:|:|        ||: |:||||||:|
Db   104668 CVGCGGNNHNNNSGGCGGCCVVGCCVGCGGNNHNNNSGGCGGCCCVGCCVGCCG 104609

Qy      312 cccacggtgaccgtcgtgttccgcacggccgtcgaggccgcgaagacgccg 371
            : |  ||:|| || | |||||:|::|:
Db   104608 GNNHNNNSGGCGGCCVGCCVGCGGNNHNNNSCGGCGGCCCCVGCCVGCCVGCGG 104549

Qy      372 tcgccgccgcgtgcgaccgcagttcaactgctacgcgccgcggcgtcggcctcccg 431
            |:|| |||  ||: ||  :| :|  |:|||
Db   104548 NNHNNNSCCGGCGCCVGCCVGCGGNNHNNNSGCCCVGCCVGCGGNNHNNNSCCGGCCVG 104489

Qy      432 ccgcacgaagccgctagcagacgcacttgtcacagggcgtgccctacaagctatcggccg 491
            :|:|: ||| : ||  || ::|     |::| |
Db   104488 CGGNNHNNNSGGCCGGCGCCGCCVGCGGNNHNNNSCCGGGCGGCCCCVGCCVGCGGGGGCGCCC 104429

Qy      492 acgacgtcttcctcaccgcgcggcgg 516
            : || :|
Db   104428 VGGCCVGCGGNNHNNNSCCGGCCGG 104404
```

RESULT 12
Q15154
ID   Q15154 standard; DNA; 4044 BP.
AC   Q15154;
DT   16-MAR-1992 (first entry)
DE   Plasmid pHAS2-MCP containing SHV-1 major capsid protein gene.
KW   Pig; recombinant vaccinia virus; pHAS2-MCP; swine herpesvirus;
KW   mad itch; ss.
OS   Pseudorabies virus.
FH   Key   Location/Qualifiers
FT   cds   10..3846
FT   /*tag= a
FT   misc_feature   1..9
FT   /*tag= b
FT   /note= "HAS2 vector and linker"
FT   misc_feature   4021..4044
FT   /*tag= c
FT   /note= "linker and HAS2 vector"
PN   J03247285-A.
PD   05-NOV-1991.
PF   27-FEB-1990; 046888.
PR   27-FEB-1990; JP-046888.
PA   (NORQ) NORINSHO.
PA   (MITU) MITSUBISHI KASEI CORP.
DR   WPI; 91-366532/50.
PT   Capsid protein of swine herpes virus type I - used for the
PT   diagnosis, prevention and treatment of Aujeszky disease
PS   Claim 3; Fig 4; 22pp; Japanese.
CC   Part of the SHV-1 MCP coding sequence is inserted into a vaccinia
CC   virus. It encodes amino acids 53 to 1330 of the major capsid
CC   protein (see R15444 for full-length sequence).
CC   See also Q15153.
SQ   Sequence 4044 BP;   577 A;   1576 C;   1327 G;   564 T;   Length 4044;

Query Match   3.38%;   Score 55.4; DB 1; Length 4044;

Best Local Similarity 46.18%;   Pred. No. 0.027;
Matches 185; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

```
Qy      262 gatcagcgcgagctggaggagacggggccggccgtgctgccctggcccacggtga 321
            || ||  |:|| ||    : ||| :   :|||||:|| | :||:|||:|
Db      651 GACCACGGCCGGCCGGCTGCGCACGGTGCTCCTGCACCACGTGCTGCCGCACACGCA 710

Qy      322 ccgtcgtgtcccgccttccgcacggcacgcctgagcgcgaagacgccgtcgccggc 381
            | || | |  |  ::|| ||||  |:  | :|| |:||  :|| ||:| |:|
Db      711 CGGCCGACGTGCCCGTGACCTACGGCGAGATGGTGATCGCCAACACCTCGTGACGGC 770

Qy      382 gctggcgcacggcacgttcaactgctacgcgccggcgtccgccgcgcacgaag 441
            ||||| | | |||||  :|  |:||  ||||:| |||| |||:|:|:
Db      771 GCTGGTGATGGGCAAGGCCGTGCAGCAACATGGACGACGTGGGCGCTACCTGCTCGGCGG 830

Qy      442 cgccgtagcagagcacttgtcacaggcgtgccctacaagctatcggccgacgacgtctt 501
            || | |  |:| |  | : | :|||  ||||: ||| |:||:|:||:||||:|
Db      831 CGAGCCGCGGCCCCGACGACGGCAAGCCCGTGGGCTGGCGGCCCCGCGGCCGCCACCTCGT 890

Qy      502 cctcacgcggactcaggcgatcggcgtcgaagtcatatccgtgtgtggccagactgc 561
            ||:|  | || | :||:|| |:  | : ||||  :|
Db      891 CGTCGGTGCGGACCGGCGGCCGCTGGTGTTCCGAGGCCCGTGCGAGAAGCGGCGTGTACCAGGCCAC 950

Qy      562 cggcgccaacatactgcttcccggccagttcatccaaattacgaggcgagcgagcgcgcatt 621
            |||:|||  :||| :|   || |||||   | :||  | | | |
Db      951 GCAGGTGCCTGCGTACCCCGTGGTGGCAACCTGGCAACCTGCACGTTCGTCATGCCCTGGCGGT 1010

Qy      622 caacaagtcgaggtccgcgcacttcgacctcatcccccgaca 662
            |   ||   || :|:| ||||| :|
Db     1011 CTTCAAGCCGGCCGGCCGCCGACCGTACGGCGCCACGCGCCGGCA 1051
```

RESULT 13
T68715
ID   T68715 standard; DNA; 15872 BP.
AC   T68715;
DT   01-SEP-1997 (first entry)
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
OS   Streptomyces venezuelae.
FH   Key   Location/Qualifiers
FT   cds   20..13912
FT   /*tag= a
FT   cds   14056..14136
FT   /*tag= b
FT   cds   14148..15827
FT   /*tag= c
PN   WO9722711-A1.
PD   26-JUN-1997.
PF   18-DEC-1996; U20119.
PR   19-DEC-1995; US-008847.
PA   (MINU) UNIV MINNESOTA.
PI   Sherman DH, Williams MD, Xue Y;
DR   WPI; 97-341701/31.
DR   P-PSDB: W19629-30 AND W00918.
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
PS   Claim 54; Fig 23; 91pp; English.
CC   Streptomyces venezuelae vep ORF1 (T68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also WI9629-30 and W00918). It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain. Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain. The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library. A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.

PR   23-DEC-1993; US-173508.
PA   (CANG) CANGENE CORP.
PI   Bartfeld D, Butler MJ, Hadary D, Jenish DL, Krieger TJ;
PI   Malek LT, Soostmeyer G, Walcyzk E;
DR   WPI; 95-24673/31.
DR   P-PSDB; R80505.
PT   Endogenous Streptomyces protease(s), opt. having impaired activity -
PT   useful in prodn. of exogenous proteins with reduced proteolytic
PT   degradation
PS   Claim 7; Fig.11; 142pp; English.
CC   Protease Tap-negative cells were transformed with a S. lividans 66
CC   genomic library and screened with APA-beta-naphthylamide to
CC   isolate colonies contg. genes (Q99365-68) for novel proteases
CC   P5-4, P5-6, P5-10 and P8-2 (R80505-08). Impaired expression of
CC   such proteases by Streptomyces hosts improves the quality, quantity
CC   and stability of exogenous gene products.
SQ   Sequence   2186 BP;   336 A;   806 C;   746 G;   298 T;

Query Match   3.3%;   Score 54.6;   DB 1;   Length 2186;
Best Local Similarity   44.2%;   Pred. No. 0.034;
Matches 225; Conservative   0; Mismatches 284; Indels 0; Gaps 0;

Qy   58 ggcaatggtacaccagagcaacggccacggccggccgccgcgcgccaacggcaa 117
             ||| | || |||| |||| ||||| || ||| ||||||| | ||||||||
Db   722 GGCCACCGTGCGGCGGGCAGCAGCCAAGGCGGCGGCGACGATCGTCACGTCGTACGACCG 781

Qy   118 gagcaacggcacgcgccgccgggccgaaccgggcaagagcaagggcacgcggcgcggcggc 177
             | | |||||| | |||| ||| || || |||||||| ||||| ||| ||||||||
Db   782 GATCGGCGGCTGATCGTCCTCCACTCCGCCAACCCGACTTCGCCAAGACCGTCGCCAAGGT 841

Qy   178 ggtggagtggaatttcgccgggggcaagaaggcatcctggcgacgacgggggcgaagaa 237
             | || | ||| ||| |||| |||||||| | || || || || ||||| ||
Db   842 GCGCGGCGTGCAGTCGCCGGTGCCACCGCGACGGCACCGCCCGCTGCCCTCGGCCGCCACCAC 901

Qy   238 cagcatccgggcgatacggtacaagatcagcgcgaggcgtggaggagcggggcgccggcc 297
             | | | || ||| | |||| | || || ||||| || || ||||| ||| ||||
Db   902 CGACACGGGCGCGCACGGTGCTCGCGGCGGTGCCCCAGGTGCTCGGCGGCGGACCTGGCGCCCAAGGCCGCCTC 961

Qy   298 cgtgctgcgcgccacggtgaccgtgacctgtgtcccggccttccgcacggccgtcga 357
             | | | ||| | | || ||| | | || || |||| ||| | || || |||
Db   962 CCGGAAGGCCGAGGGCCAGAGGCCCAGCACCCTGGAGTCGCTCGCCAGTGGGACCTGCCCGCCCATCAA 1021

Qy   358 ggccgaagagacgccgtcgcctgccgctgcacggccagttcaactgtacgccgccgg 417
             | | ||| | |||| || || || | || || | ||| ||||| | ||||||||
Db   1022 GGCGGACAAGGCGCACGAGAAGTCGTCGGCAGCAGGAAGGTGACCGTCGCCGTCATCGA 1081

Qy   418 cgtcggcctcccgccgccgcacgaagcgcgtagcagagcacttgtcacagggggtgccta 477
             | | || || ||||| || || | |||| | | || |||| | ||||| | | || |
Db   1082 CACCGGCGTCGACGACCACGCACCCCGACATCGCCCGGACAACTTCGACCGGACGGGTCCGT 1141

Qy   478 caagctatcggccgacgacgtcttcctcaccgccggcggacctcaggcgatcggaagtcat 537
             ||| | ||| ||| | | | | | ||||| ||| || | | |||| || || |||
Db   1142 CAACTGTGTGCGGGGCAAGCCGGACACCGCCCACGGCGGCCTGCGCGGCGGAGCGGCGGCGGGA 1201

Qy   538 aatcccggtgctggccagactgccgcg 566
             | ||||| | | | || | || ||
Db   1202 GAGCCCGCACACGCACCCACGTGGCCGGGG 1230

RESULT 15
ID   V84066 standard; DNA; 2185 BP.
AC   V84066;
DT   19-MAR-1999 (first entry)
DE   Clone P5-4 encoding a protease.
KW   Tripeptidyl aminopeptidase; TAP; N-terminal cleavage; protein production;
KW   GM-CSF; interleukin-3; IL-3; IL-6; EPO; tumour necrosis factor; TNF;
KW   SCF; IL-7; IL-2; P5-4; ds.
OS   Streptomyces lividans
FH   Key   Location/Qualifiers
FH
FT   CDS   531. .2069
FT   /*tag= a

SQ   Sequence   15872 BP;   2093 A;   6313 C;   5503 G;   1963 T;

Query Match   3.3%;   Score 55;   DB 1;   Length 15872;
Best Local Similarity   44.8%;   Pred. No. 0.042;
Matches 290; Conservative   0; Mismatches 355; Indels 2; Gaps 2;

Qy   73 gagcaacggcacggccgagggccgccgccgccaacggcaagacgaacgggcacgc 132
             ||| | || |||| |||| ||||| || ||| ||||||| | ||||||||
Db   1264 GGGCCGGCCACCAAGCCCACGTCGTCCTCAGCGAAGGCGGCCCCAGGCGGCGAGCGAGCC 1323

Qy   133 cgccgccgcgaacggcaagagcaacgggcacgggccgcgggcgcggtggagtggaatttt 192
             ||| || || || || | | |||| ||| || || | | ||| | || ||||
Db   1324 CGGCATCGATGAGGAGACCCGTCGACAGCGGGGGCCGCGACTGCCCTTCGTCGTCACCGG 1383

Qy   193 cgccgggcgaaggacggcaaggacggcgaaggacagcagcatccgggc-ga 251
             |||| ||| | | | || | | || | ||| ||||| |||||| ||
Db   1384 CCGGCGGCGGAGGCCCTGCGCGCGCCCAGGCCCGCGCCCTGCGCAGCAGGCCGTGAAGCGGA 1443

Qy   252 tacggtacaagatcagcgcgagcgtggagagcggggcccgtgctgcgctgg 311
             ||| ||| | || || ||| ||| |||| ||| ||| || ||| ||| |
Db   1444 CCCGGAGCTCGCGCCCGCCACTCGCTCGTCGTCACCACCCGTACGGTCTTCAC 1503

Qy   312 cccacggtgaccgtccgtttccgcacggccttccgcacggccgtcgaggccgaagacgcg 371
             ||||| | | | | || |||| ||| || ||| || | | ||||| |||| | |
Db   1504 GCACCGGTCGTGCGTCTTCCGCGGCGCCCCTCC-TCGACGGCCTCGGCGCCC 1562

Qy   372 tcgccgccggcgctgcgcaccggcagttcaactgtcacctgtcaagctatcggccg 431
             ||| | || || || |||| || ||| ||||| |||| ||| | || || ||
Db   1563 TCGCCGCCGGGACGCCCGCCCGGCACCCCCGGGGCCTCG 1622

Qy   432 ccgcacgaagcgccgtagcagagacagttgtcacagggtgccctacaagctacggccg 491
             ||| ||| || || || ||||| || || ||||| ||| ||||| |||| ||||||
Db   1623 CCGTTCGTTCAGCGGCAGGTGCCCAGGTGCCCCAGGTCCCCTACGCCGGAGTTGTACGCCG 1682

Qy   492 acgacgtcttcctcacgcgcggcgaactcaggcggatcgaagtcataatcccggtgctgg 551
             | | | | ||||| |||| ||| || | || ||| | || | || ||||||| ||| |
Db   1683 CCCACCCCGCCTTCGCGACGGCCTTCGCCGGCGTCGCCGCCGCGAACTGGACCCCCTCG 1742

Qy   552 cccagacgctgcggccaacatactgcttcccggccagctatccaaattacgaggcgc 611
             ||||| ||| ||| | | | | || ||||| || || | | | || ||| ||
Db   1743 ACCGGCCCCTCGCTCGCGAACTGCTGTCCGGCGGCGACACCCTGCAGGCCACGTCCACACAC 1802

Qy   612 gagcggcattcaacaagctggagtccggcacttcgacctcatcccgacaagggtggg 671
             || ||||| ||||| ||| | || || || | ||| | | || || || || || |
Db   1803 AGCCGCGGCGCTTCGCCGTGGAGGTCGCCGTGCCCCTCCACCCGCGTCCGTCAGTCCTGGGGCGTCA 1862

Qy   672 agatcgacatcgactcgctggaatccatcgcgacaagaacaccacc 718
             || ||| | || | | || || ||| || || | || |||
Db   1863 CGCCGGCGTCGACCTGCTCCGCGGCCACTCGCTCGCGAGATCAGCGCCCCC 1909

RESULT 14
Q99365
ID   Q99365 standard; DNA; 2186 BP.
AC   Q99365;
DT   04-DEC-1995 (first entry)
DE   S. lividans protease P5-4 gene.
KW   Protease; metalloendoproteinase; tripeptidyl aminopeptidase;
KW   protease-deficiency; protein secretion; ds.
OS   Streptomyces lividans
FH   Key   Location/Qualifiers
FH
FT   cds   531. .2069
FT   /*tag= a
FT   signal_peptide 531. .901
FT   /*tag= b
FT   /note= "includes propeptide sequence"
FT   mat_peptide   902. .2063
FT   /*tag= c
FT   /product= P5-4.
PN   WO9517512-A.
PD   29-JUN-1995.
PF   22-DEC-1994; U14772.

```
FT   sig_peptide      /product= protease
FT                    531..902
FT                    /*tag= b
FT   mat_peptide      903..2069
FT                    /*tag= c
PN   US5856166-A.
PD   05-JAN-1999.
PF   24-JUN-1994; 265310.
PR   24-JUN-1994; US-265310.
PR   23-DEC-1993; US-173508.
PA   (CANG-) CANGENE CORP.
PI   Bartfeld D, Butler MJ, Garven S, Hadary D, Jenish DL,
PI   Krieger TJ, Krygsman P, Malek LT, Soostmeyer G, Walczyk E;
DR   WPI: 99-105117/09.
DR   P-PSDB; W87797.
PT   Streptomyces tri:peptidyl aminopeptidase - useful for removing
PT   N-terminal pro-peptide from secreted proteins
PS   Example 20; Fig 20A-C; 83pp; English.
CC   The present sequence encodes a protease. The specification also
CC   describes a tripeptidyl aminopeptidase (TAP) of Streptomyces.
CC   The aminopeptidase is endogenous to Streptomyces and cleaves an
CC   N-terminal sequence of X-Pro-Y, where x is an aliphatic or hydroxy
CC   amino acid and y is an aliphatic, hydroxy or sulphur-containing
CC   amino acid. The TAP of Streptomyces are useful in the production
CC   of proteins, such as GM-CSF, interleukin-3 (IL-3), IL-6, EPO,
CC   tumour necrosis factor (TNF) SCF, IL-7 and IL-2.
SQ   Sequence 2185 BP;   336 A;   806 C;   745 G;   298 T;
```

```
Query Match             3.3%;  Score 54.6; DB 1; Length 2185;
Best Local Similarity  44.2%;  Pred. No. 0.034;
  Matches 225; Conservative    0; Mismatches 284; Indels    0; Gaps    0;
```

```
Qy     58 ggcaatggtacaccagagacaacggccacggccgagcgccgccgccaacggcaa 117
          ||| | ||  |  | ||   |   | ||| |  | || | ||||| |  | |||
Db    722 GGCCACCGTGCGGCGGGGCATAGCCAAGGCGGGCGGCACGATCGTCACGTCGTACGACCG 781

Qy    118 gagcaacggcacgccgccgcccgcgaacgcaagagagcaacggcacgcggcggcggcggc 177
          | |   ||  ||  | || || | | |   |  | | |  |  | |     |  |  | 
Db    782 GATCGGGCGTGATCGTCGTCCACTCGCCAACCCCGACTTCGCCAAGACCGTGCCAAGGT 841

Qy    178 ggtggagtggaatttcgccgggggcaaggacggcatcctggcgacgacggggcgaaga 237
           || |  ||  | |  |   |  || |   || || | |  | | | || | | | |||
Db    842 GCGCGGCGTGCAGTCGGCGTGCCACCCGCCGCCGTGCCTCGGCCGCCGCCACCAC 901

Qy    238 cagcatccgggcgatacggtacaagatcagcgcgagcgtggagagcgggccgcggcc 297
          || |  ||  || | | ||   || ||  | | |  | |  | |  | | |||  || 
Db    902 CGACACGGGCGCGCGCCAGGTGCTCGGCGGCGAGGACCTGGCCGCCGACCTCGGCCCTC 961

Qy    298 cgtgctgccgctggccacgtgaccgtccgtgttccggccttccgcacggccgtcga 357
          || | || | | | || | | |   | | |  | |  || |  | |  | | | |||
Db    962 CGGAAGGCCGAGGCGCCAGGACCCGCTGGAGTCGCTCCAGTGGGACCTCCCCGCCATCAA 1021

Qy    358 ggccgaagacgccgtcgccgccgcggcctgcgcaccggccagttcaactgctacgccgcgg 417
          || | | | |  | | |  | | |  | || || |  || || | || | || |  ||||
Db   1022 GGCCGACAAGGCCGCACCAGAGAGTGCTGGCCAGCAGGAGGTGACCGTGCCCGTCATCGA 1081

Qy    418 cgtcggcctccccgccgcacgaagcgccgtagcagagcacttgtcacagggcgtgcccta 477
          || | ||  ||  |  ||  | | | ||| ||  | | |||  |  |   |  || | |||
Db   1082 CACCGGCGTCGACGACACCCACCGGACATCGCCCGACTTCGACCGGCCAGGCGTCCGT 1141

Qy    478 caagctatcggccgacgacgtcttcctcaccgccgcggaactcaggcgatcgaagtcat 537
          || |  | || | | | ||   | || | | ||  | ||   || | | |  ||||||
Db   1142 CAACTGTGTGGCGGGCAAGCCGACACCCGACGGGCGGCCTGGCGGGCCAGCCGCGGCGGA 1201

Qy    538 aatcccggtgctgcccagactgccggcg 566
          || ||  | | |  | |    |||||||
Db   1202 GAGCCCGCACGGCACCCACGTGGCCGGGG 1230
```

```
Search completed: October 1, 1999, 15:35:15
Job time: 6053 sec
```

us-09-026-400-1.rng

OM nucleic - nucleic search, using sw model

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: October 1, 1999, 13:50:36 ; Search time 1096.08 seconds
                                          (without alignments)
                                          2987.378 Million cell updates/sec

Title: US-09-026-400-1
Perfect score: 1660
Sequence: 1 attgactagctagttcattc........aaaaaaaaaaaaaaaaa 1660

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues
Database :   EST:*

```
 1: em_est1:*           20: gb_est1:*           40: gb_est21:*
 2: em_est2:*           21: gb_est2:*           41: gb_est22:*
 3: em_est3:*           22: gb_est3:*           42: gb_est23:*
 4: em_est4:*           23: gb_est4:*           43: gb_est24:*
 5: em_est5:*           24: gb_est5:*           44: gb_est25:*
 6: em_est6:*           25: gb_est6:*           45: gb_est26:*
 7: em_est7:*           26: gb_est7:*           46: gb_est27:*
 8: em_est8:*           27: gb_est8:*           47: gb_est28:*
 9: em_est9:*           28: gb_est9:*           48: gb_est29:*
10: em_est10:*          29: gb_est10:*          49: gb_est30:*
11: em_est11:*          30: gb_est11:*          50: gb_est31:*
12: em_est12:*          31: gb_est12:*          51: gb_est32:*
13: em_est13:*          32: gb_est13:*          52: em_est20:*
14: em_est14:*          33: gb_est14:*          53: em_est21:*
15: em_est15:*          34: gb_est15:*          54: em_est22:*
16: em_est16:*          35: gb_est16:*          55: em_est23:*
17: em_est17:*          36: gb_est17:*          56: em_est24:*
18: em_est18:*          37: gb_est18:*          57: em_est25:*
19: em_est19:*          38: gb_est19:*          58: em_est26:*
                        39: gb_est20:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	372.6	22.4	571	49	AI657244	AI657244 486092E01
c	2	301.8	18.2	594	49	AI619119	AI619119 486073F11
	3	262.6	15.8	459	21	D48521	D48521 RICS14756A
	4	240.2	14.5	410	20	D25143	D25143 RICR3343A R
	5	177	10.7	464	49	AU056779	AU056779 AU056779
	6	150	9.0	446	20	Z38084	Z38084 ATTS4224 Ve
	7	136.8	8.2	470	23	T76370	T76370 11148 Lambd
	8	122.8	7.4	343	20	D23043	D23043 RICC2086A R
	9	117.8	7.1	352	23	T13684	T13684 1849 Lambda
	10	106.6	6.4	719	40	AA980369	AA980369 ua52c12.r
c	11	102.4	6.2	283	49	AI621525	AI621525 486092E01
	12	95.6	5.8	512	34	AA511748	AA511748 vj29f11.r
c	13	90.6	5.5	405	20	Z38047	Z38047 ATTS4207 Ve
	14	86.8	5.2	347	23	T41757	T41757 10338 Lambd
	15	85	5.1	415	48	AI597379	AI597379 vj29f11.y
	16	84.8	5.1	757	40	C91185	C91185 Dict
	17	77.8	4.7	559	33	AA394723	AA394723 26506 Lam
	18	77.6	4.7	536	47	AI487927	AI487927 EST246249
	19	75	4.5	801	42	AI116859	AI116859 ue29d01.y
	20	74	4.5	729	41	AI048570	AI048570 ud61e01.y
	21	66.6	4.0	521	42	AI116355	AI116355 uf02b01.y
	22	63.6	3.8	528	25	N97159	N97159 22338 Lambd
	23	63.4	3.8	549	26	W43288	W43288 22681 Lambd
	24	60.2	3.6	664	41	AI055475	AI055475 coau0004B
	25	59.8	3.6	699	45	AI386363	AI386363 mo05a03.Y
	26	57.8	3.5	591	28	AA087374	AA087374 mn96b03.r
	27	57.4	3.5	288	36	AA650709	AA650709 30823 Lam
	28	57	3.4	437	20	T53940	T53940 yb85b03.r1
	29	56	3.4	248	36	C72041	C72041 C72041 Rice
	30	53.2	3.2	644	40	AA979730	AA979730 MEST1-B1.
	31	53.2	3.2	618	40	AA979961	AA979961 MEST4-C10
	32	53.2	3.2	621	40	AA980005	AA980005 MEST4-C5.
	33	52.4	3.2	232	20	D41250	D41250 RICS3617A R
	34	52.2	3.1	870	48	AI562691	AI562691 TENS2678
c	35	51.8	3.1	664	48	AI550776	AI550776 mn96b03.Y
	36	51.6	3.1	502	36	AA395846	AA395846 25834 Lam
	37	50.4	3.0	719	38	AA752790	AA752790 97AS0372
	38	50.4	3.0	779	48	AI562604	AI562604 TENS2761
	39	50.4	3.0	557	48	AI601030	AI601030 486096D01
c	40	49.2	3.0	623	31	AA292171	AA292171 zt50d08.r
c	41	48.8	2.9	328	42	AI138530	AI138530 qd80f01.x
c	42	48.8	2.9	650	46	AI442831	AI442831 sa27c08.x
c	43	48.8	2.9	198	49	AI626616	AI626616 fc05e10.x
c	44	48.6	2.9	480	42	AI132468	AI132468 ue96e10.x
	45	48.4	2.9	461	49	AU056780	AU056780 AU056780

ALIGNMENTS

RESULT 1
AI657244
LOCUS   AI657244   571 bp   mRNA   EST   05-MAY-1999
DEFINITION 486092E01.y1 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI657244
NID   g4753339
VERSION   AI657244.1 GI:4753339

```
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1 (bases 1 to 571)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189324.

            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486092 row: E column: 01.
FEATURES    Location/Qualifiers
   source   1..571
            /organism="Zea mays"
            /cultivar="B73"
            /db_xref="taxon:4577"
            /clone_lib="486 - leaf primordia cDNA library from Hake
            lab"
            /tissue_type="leaf primordia"
            /dev_stage="P7-P11 leaf"
            /lab_host="E.coli XL1-Blue MFR'"
            /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
            library."
BASE COUNT    108 a     198 c     178 g      87 t
ORIGIN

Query Match
Best Local Similarity     22.4%;        Score 372.6;   DB 49;   Length 571;
                          80.3%;        Pred. No. 4.9e-55;
Matches 462;   Conservative    0;   Mismatches 109;   Indels 4;   Gaps 2;

Qy   257 tacaagatcagcgagcgtggagagagcggcggcccgtgctgccgtgccccac 316
             ||| || |||| |||   |   ||||||  |||| |||  || |
Db     1 TTCAAGATCAGCGGCGACCCTGACGCCGGCGCCCGGCCCGTGCCCGGCGCCAC 60

Qy   317 ggtgaccgtccgtcgttcccggccttccgcacggccgtcgaggccgaagacgccgtcgcc 376
             || |||  | |||| |||| ||| |   ||| ||||||  | |
Db    61 GGGGACCCCTCGGTGTTCCCGGGCCTTCCGCACCGCGAGCCGGAGCGCCCGTGGCC 120

Qy   377 gccggcgtgcgcaccggccagttcaactgctacgcccgccgggtcggcctcccgccgca 436
             |||||| ||  ||||| ||  ||||| |||| |||| ||  ||||| |||| || ||
Db   121 GCCGGCGTCCTCGACCGGCAAGTTCAACTGCTACCCCCGCCGGCGGTCGGCCTCCCGAAGCC 180

Qy   437 cgaagcgccgtagcagcacttgtcacaggcgtgccctacaaagctatcggccgacgac 496
             || |||||  |  ||| ||| | |||||| |||||||||| |||||| ||||| || ||
Db   181 CGCCGGTCGTCTCGGCGGAGCACCGTGCGAGCACCTTCGAGCCGGACTGTCGAGCGACGAC 240

Qy   497 gtcttcctcaccgccggcggaactcaggcgatcgagtcataatcccgtgctgcccag 556
             || ||||| ||| |||| |||| |||| || |  || || | |  ||| |  | ||||
Db   241 ATCTTCGTCACCGCCGGAGGCACGCAAGCCATCGAGGTGGTCTCAGTCCTCGCGCAG 300

Qy   557 actgccggccaacatactgcttcccggccaggctatccaaattacgaggcgcgagcg 616
             | |    ||||  ||| ||| |||| |||| || | ||| ||||  |  || | ||
Db   301 ---CCGGGCACCAACATACTGCTCCGGAGCCTATCCGAACTACGAGCGCCGCGCCGCA 357

Qy   617 gcattcaacaagctgagtccggcacttgacctcgacctcatcccgacaagggtgggagatc 676
             |||| |||||||| | | | ||  | | ||| ||| | | | ||
Db   358 GGGCTGCACAACCTGGAGTTCGCCGGTTCAATCGATCCCGTTCCCGAGAGGGGTGGGAGATT 417

Qy   677 gacatgactcgtgctgaatcatcgccgacaagacaccaccacggtgtcatcataaac 736
             ||||| |||||  || |||  ||| || |||| |   || || |   || |||||  ||
Db   418 GACATCGACGGTCTCGAGTCGATCGCCGACGAGCAAGAACACCACCGCCATGGTCATCATAAAC 477
```

```
Qy   737 ccaaacaatccgtgcggcagcgttactcctacgacccatctggccaaggtcggcgaggtg 796
             || |||| || ||| |||||| |   || ||||  |  ||| |||||||| |||| |
Db   478 CCCACAACCCTTGCGGGAGTGTCTCACACCCGAGAGCATTTGGCCAAGGTCGCGGGA-GTG 536

Qy   797 gcaaggaagctcggaatattggtgatgctgacga 831
             | ||||||||| |||  || |||| |||| ||
Db   537 GCAAGGAAGCTTGGAATACTGGTCATCGCTGATGA 571


RESULT 2
AI619119/c
LOCUS       AI619119         594 bp    mRNA              EST       21-APR-1999
DEFINITION  486073F11.x2 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI619119
NID         g4628245
VERSION     AI619119.1   GI:462824 5
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1 (bases 1 to 594)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     On Mar 10, 1998 this sequence version replaced gi:2948200.

            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486073  row: F  column: 11.
FEATURES    Location/Qualifiers
   source   1..594
            /organism="Zea mays"
            /cultivar="B73"
            /db_xref="taxon:4577"
            /clone_lib="486 - leaf primordia cDNA library from Hake
            lab"
            /tissue_type="leaf primordia"
            /dev_stage="P7-P11 leaf"
            /lab_host="E.coli XL1-Blue MFR'"
            /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
            library."
BASE COUNT    172 a     113 c     112 g     197 t
ORIGIN

Query Match
Best Local Similarity      18.2%;        Score 301.8;   DB 49;   Length 594;
                           78.8%;        Pred. No. 6.7e-43;
Matches 360;   Conservative    0;   Mismatches 97;   Indels 0;   Gaps 0;

Qy   996 aaactaagatctctacgtctattacgaattacctaatgtctcaacggaccagcaacct 1055
             |||| ||||| |   |||||| ||| |  |||| ||| || |||| || ||||| ||
Db   589 AAACCAAGATCATTGCATCGATCACAAACTTCCTTAACGTTTCAACAGATCCGCACTT 530

Qy  1056 tcgttcaggaagctcttcctaaaattcttgagaacaacaaaagcagatttcttaagagga 1115
             | |||||| ||| |||| | | ||||||| |||||||||| || |||| ||| ||||||
Db   529 TTGTTCAGGGAGCGGCTTCCCCATATTCTTAAGAACAACAACAAAGGAAGAATTTTTCAAGAGGA 470

Qy  1116 ttattggtctactaaaggaatcatcagaatagttatagggaaataaaggaaaacaaat 1175
             | |||||| |  || ||| ||||||| ||||| || | | |||| ||| || | |||
Db   469 TCATTGGTCGTGCTAGCGGAAACATCAAGAATATGTTTCAGCGGAATAAAAGGACATCAAGT 410

Qy  1176 atattacgtgtctcacaagccagaaggagatcgatgtttgtaatggtcaaacaacttac 1235
             | || |||| | ||| ||||||||| || | |||||||||| |||| |||| ||  |
Db   409 GCATCACATGCCCTCCCAAGCCAGAGAGGTTCCATGTTTGTCATGGTGAAACTAAATTTGT 350
```

```
Qy   1236  atcttttggaggagatccatgacgacatagatttttgtgcaagctcgcaaaggaagaat  1295
                ||||||||||||| |||||||||||  ||  |  ||||||  |||| ||||| ||
Db    349  ATCTTTGGAGAGCATCCATGATGATATTGATTTTTGTTGTTGCAAGCTGGCAAAGAAGAGT  290

Qy   1296  cagtaaatttatgtccaggagtgttcttggaatgagaaaattggtccgtattacttttg  1355
                || ||| |  |||||| ||| |||| ||||||| || | || |||| |||||||||
Db    289  CCGTGATTTTGTGTCCAGGAGTGTTTTTGGAATGAAAACTGATCCGTATCCACTTTCG  230

Qy   1356  cctgcgttccatcttctcttcaagatggactcgaaaggtcaaatcattctgtcaaagga  1415
                ||  || ||||||||  ||  ||||||||||||||| ||  ||| |||||||||||||
Db    229  CCATTGATTCATCTTCTTCTTCTTGATGTGTCTTGAGAGGTCTGAAATCTTTCTGCCAAGGC  170

Qy   1416  acaagaagaagaattctataaatgttggttagttgta  1452
                ||||||||||| |  | |  || ||||  |||  ||
Db    169  ATAAGAAGAAGAATTTGCTTAATGGCCATTAACTATA  133
```

RESULT   3
D48521
LOCUS   D48521   459 bp   mRNA   EST   02-AUG-1995
DEFINITION RICS14756A Rice green shoot Oryza sativa cDNA, mRNA sequence.
ACCESSION   D48521
NID   g702230
VERSION   D48521.1 GI:702230
KEYWORDS   EST.
SOURCE   Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 459)
AUTHORS   Sasaki,T., Miyao,A. and Yamamoto,K.
TITLE   Rice cDNA from callus 1995
JOURNAL   Unpublished (1995)
COMMENT   Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            High quality sequence stop: 405.
            Insert Length: 852   Std Error: 0.00
FEATURES   Location/Qualifiers
     source   1..459
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone_lib="Rice green shoot"
                     /note="Green shoot (8 days old)"
BASE COUNT   134 a   86 c   105 g   132 t   2 others
ORIGIN

Query Match   15.8%; Score 262.6; DB 21; Length 459;
Best Local Similarity   74.4%; Pred. No. 3.5e-36;
Matches 343; Conservative   0; Mismatches 116; Indels   2; Gaps   1;

```
Qy    672  agatcgacatcgactcgctcgaatccatcgccgacaagaacaccaccgcgatggtcatca  731
                |||||| | ||||| ||||| |||  ||||  |||||| |  | ||| ||| |||||
Db      1  AGATTGACCTTAACTCCCTAGAATCTATTGCGACACTACTCGGATAGTCATCA  60

Qy    732  taaaccaacaatccgtgcgcagcgtttactcctacgaccatctggccaggtcgcgg  791
                || ||||  | | |||| |||||||  || || ||||||| ||| || |||| ||
Db     61  TAAATCCCAATAATCATGCGGGAATGTGTCAACTTACGACGATTATCCAAGGTGGCAG  120

Qy    792  aggtggcaggaagtccggaatattggtgatcgctgacgaggtttacggcaaactggttc  851
                |||||||||||| || |  ||  | ||||| ||||||||  || |||| |||| |||
Db    121  AGGTAGCAAGGAAGCTTGGGNTATTGGTAATTACTGATGAGGTGTATGGGTAATTTGGTTT  180
```

```
Qy    852  tgggcagcgcccgtttatccgcgatgggcgtcttggcacattgccccgtcttgtcca  911
                ||||| || |||||| ||||| |||| ||  | ||||  ||||||| ||| |||||||
Db    181  TTGGGAGTTCCCCATTGTCCCCAATGGTTGCTTTGGGCACATCGCTACCAATATTAACCA  240

Qy    912  ttggatctctgtccaagtctggaatagtgccttggtgctgactggatggtggcggtgt  971
                | |||| || ||||| |||  ||||||||| ||| || ||||| ||| ||| |||| |
Db    241  TAGGATCGCTATCAAAGAGGTGGATAGTGCCGGATGGCCGGATGGGCGACTTGGTTGGGTAGCAATAT  300

Qy    972  acgaccccacaaagatttagagaaaactaagactctctagtcattacgaattacctta  1031
                |  ||| | || || ||| ||||||||| || || |||| ||||| || |  |||| |
Db    301  GTGACCCCAAGAGACTCTCACAAGAGAAACCAAGNTGCAACATTAATTACTAATTCCTTA  360

Qy   1032  atgtccaacgaccaggaaccttcgttcaggaagctcttcctaaaaatttcttgagaaca  1091
                ||||| || || || |||| ||| || || ||||| | || ||||  | || ||||||
Db    361  ATGTTTCACTGATCCAGCAACTTTCATTCAGGGAGGTAATTGATTTGCTTACG  420

Qy   1092  caaaagcagatttcttaagaggattattggtctactaaag  1132
                ||  || ||||  || |||||  |  || | ||  ||||
Db    421  C--AAGGAGATTCTTAAGGGGGTAATTGATTTGCTTACG  459
```

RESULT   4
D25143
LOCUS   D25143   410 bp   mRNA   EST   20-JUL-1998
DEFINITION RICR3343A Rice root Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D25143
NID   g428987
VERSION   D25143.1 GI:428987
KEYWORDS   EST.
SOURCE   Oryza sativa.
ORGANISM   Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1 (bases 1 to 410)
AUTHORS   Minobe,Y. and Sasaki,T.
TITLE   Rice cDNA from root
JOURNAL   Unpublished (1995)
COMMENT   Contact: Yuzo Minobe
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: minobe@rtcs0.riken.go.jp
            PROJECT ='RGP',
FEATURES   Location/Qualifiers
     source   1..410
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="R"
                     /clone_lib="Rice root"
                     /note="Prepared from seedling root. "
BASE COUNT   125 a   69 c   97 g   119 t
ORIGIN

Query Match   14.5%; Score 240.2; DB 20; Length 410;
Best Local Similarity   75.8%; Pred. No. 2.4e-32;
Matches 310; Conservative   0; Mismatches 98; Indels   1; Gaps   1;

```
Qy    788  gcggaggtggcaggaagctcggaatattggtgatcgctgacgaggtttacggccaaactg  847
                || || || || |||| | |||  ||||| | |||| || ||| |||| || || ||
Db      2  GCAGAGGTAGCAAGGAAGCTTGGGGATATTGGTAATTACTGATGAGGTGTATGGGTAATTG  61

Qy    848  gttctggcagcgccccgtttatccgcgatgggcgtcttggcacattgccccgtcttg  907
                |||| || || |  |||||| |||| ||||| ||  | ||| |||| || |||  |||
Db     62  GTTTTTGGGAGTTCCCCATTGTCCCCAATGGTTGCTTTGGGCACATCGCTACCAATATTA  121
```

```
Qy  908 tccattggatcctgtccaagtcgtggatagtagtagtgcgacttggatgggtggcg  967
        |||||  ||||  ||  |  |  ||||||  || || || |||||||||| ||| ||
Db  122 ACCATAGGATGCTATCAAAGAGGTGGATAGTGCCGGATGGCGACTTGGTTGGGTAGCA  181

Qy  968 gtgtacgaccccacaaagattttagagaaactaagatctctacgtctattacgaattac 1027
        | | ||||||||  || || |  | || ||  || |||||  ||| |||| || || ||
Db  182 ATATGTGACCCCAAGAGACTCTACAGAGAACAAGAAAACCAAGATTGCAACATTAATTACTAATTTC  241

Qy 1028 cttaatgtctcaacggacccagcaaccttcgttcaggaagctcttcctaaaattcttgag 1087
        ||||||||  ||  || ||||||||| || | ||||| |||| |||| ||| ||||| |
Db  242 CTTAATGTTCACTGATCCAGCAACTTCATTCAGGGAGCTCTACCGAATATTCTTAAG  301

Qy 1088 aacacaaagcagattcttaagaggattattggtctac-taaaggaatcatcagagat 1146
        || || ||| || ||||||||||| || ||||| ||| |||| |||||||||| ||
Db  302 AATACCAAGGAAGAATTCTTTAAGAGGATAAATTGATTTGCTTACGGGAAACATCAGATAT  361

Qy 1147 atgttataggg aataaagagaaacaaatattacgtgtcctcacaag 1195
        | | ||| ||| |||||| || || ||||| ||| |||||||||||
Db  362 TTGCTATAGAGGGATAAAGGGTATTAAATGCATCACTTGTCCTCACAAG  410
```

RESULT 5
AU056779

```
LOCUS       AU056779      464 bp    mRNA              EST       29-APR-1999
DEFINITION  AU056779 Oryza sativa mature leaf Nipponbare Oryza sativa cDNA
            clone S20870_1A, mRNA sequence.
ACCESSION   AU056779
NID         g4715663
VERSION     AU056779.1  GI:4715663
KEYWORDS    EST.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 464)
AUTHORS     Yamamoto,K. and Sasaki,T.
TITLE       Rice cDNA from mature leaf
JOURNAL     Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3187002.
            Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT='RGP'
FEATURES             Location/Qualifiers
     source          1..464
                     /organism="Oryza sativa"
                     /strain="Nipponbare"
                     /db_xref="taxon:4530"
                     /map="21q"
                     /clone="S20870_1A"
                     /clone_lib="Oryza sativa mature leaf Nipponbare"
                     /tissue_type="mature leaf"
BASE COUNT    117 a    105 g    132 t    109 c    1 others
ORIGIN

Query Match             10.7%;       Score 177;  DB 49;  Length 464;
Best Local Similarity   62.7%;       Pred. No. 1.6e-21;
Matches   292;  Conservative  0;  Mismatches 171;  Indels  3;  Gaps 0;

Qy  403 ctgctacgccgccgcgtcggcctccccgccgcaagcgcgcctagcgagagcacttgtc  462
        | ||||| || || |  |  |||||| ||| | |||| | || |||  | || || |||
Db    1 CGGCTACNCNCTCCCCCGCCACCAGCCTCCCTGCTCGCCGAGCTATTGCGAGAGTACCTATC  60
```

```
Qy  463 acagggggtgccctacaagctatcggccgacgacgtcttcctcaccgccggcggaactca  522
        |||| |    |  |||||||| | ||  | | || |  |||||| |  |||  |  ||| |
Db   61 CTCTGATCTTCCTTACAAGCTTTGCACAGATGATATTTCCTCACATCTGGAGTACCCA  120

Qy  523 ggcgatgaagtcatatccgtgctggcccagactgccggcgccaacatactgcttcc  582
        |||  |  | | | |  |||| || || ||||||| ||||||||  ||| ||||||
Db  121 AGCAATCGAGATTGTTATGTCTGTTTTTGGCCA---ACCAGGTGCCAATATATTGCTTCC  177

Qy  583 ccggccaggctatccaaattacgagcgcggagcggcattcaacaagctggaggtccggca  642
        || | ||| | | | | | ||   ||| | |  || ||| |||  | ||||| |  ||||
Db  178 AAAGCCCGGTACCCAAAACATGAAGCACATGGGTGTTCCACAGGATGGAAGTGCGGCT  237

Qy  643 cttcgacctcatccccgacaagggggtgggagatcggacatcgactcgctggaatccatcgc  702
        || ||| ||| ||  | | ||||| ||| |||| || ||||  ||| |||| |||||||| |
Db  238 CTATGATCTTGTTCCAGAGAGGATGGGAGATTAANTGTTGAAGCTGTTGAAGCTTTAGC  297

Qy  703 cgacaagaacaccaccgcgtgtcatcataaaccaaacaatccgtgccgagcgttta  762
        || |||||| || ||  |||||||||||| |||||||||| |||| | | |||| |
Db  298 AGATGAGAATACTGTTGCAATAGTGATTACTAACCCCAATAACCCTTGTGGGTAATGTGTA  357

Qy  763 ctcctacgacatctgcgcaggtcgcggaggtcgcggaggaagctcggaatattggtgat  822
        || |||| | ||||| |||| | || |||| | ||| ||| |||  || || |||| ||
Db  358 CACTTATGCAGCATCTGTCCAAGATTGCAGACAGCAAGCAAGTTCGTCTGTTGATTCAT  417

Qy  823 cgctgacgaggtttacggcaaactggttctgggcggcagcgcccccgttt  868
        |||||| ||| | |||||  ||||||| | |  | ||||| | ||| ||||
Db  418 TGCTGATGAAGTATAGTGGTCACCTTGTTTATGGAAGCACTCCTTT  463
```

RESULT 6
Z38084

```
LOCUS       Z38084        446 bp    mRNA                      EST    24-OCT-1994
DEFINITION  ATTS4224 Versailles-VB Arabidopsis thaliana cDNA clone VBVMA10 5',
            mRNA sequence.
ACCESSION   Z38084
NID         g558161
VERSION     Z38084.1  GI:558161
KEYWORDS    EST.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 446)
AUTHORS     The Arabidopsis thaliana transcribed genome: the GDR cDNA program
TITLE
JOURNAL     Unpublished (1996)
COMMENT     Contact: Desprez T., Amselem J., Chiapello H., Rouze P., Caboche
            M., Hofte H.
            INRA Versailles
            Laboratoire de Biologie Cellulaire
            Route de Saint-Cyr,78026 Versailles Cedex,France
            Email: thierry@versailles.inra.fr.
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Arabidopsis thaliana"
                     /strain="ecotype Columbia"
                     /db_xref="taxon:3702"
                     /clone="VBVMA10"
                     /clone_lib="Versailles-VB"
                     /note="vector: pBluescript; tissue=whole seedlings of A.
                     thaliana ecotype Columbia; clone_library=Versailles-VB;
                     Cloning vector: pBluescript (Stratagene); Physiological
                     conditions: in vitro-grown etiolated seedlings,5 days
                     old."
BASE COUNT    100 a     81 c    111 g    146 t
ORIGIN

Query Match             9.0%;        Score 150;  DB 20;  Length 446;
Best Local Similarity   60.8%;       Pred. No. 7e-17;
```

Matches 271; Conservative 0; Mismatches 173; Indels 2; Gaps 2;

```
Qy    629 ctggaggtccgcacttcgacctcatcccgacaagggtgggagatcgacatcgactcg 688
          ||||| || || |||| || ||| | |||||||| |||||||||||||| | |
Db      1 CTNGAANTTCGTACGTCGATCTTCTTCCGGAAAATGGATGGGAGATCGATCTTNATGNN 60

Qy    689 ctggaatccatcgccgacaagaacaccaccgcgatg-gtcatcataaacccaaacaatcc 747
          || |||||| || || || | | || | ||| ||| || |||||||| |||| ||
Db     61 GTCGAGGCTCTTGCAGACGAAAACACGGTTGCTTTANTTGTNATAAACCCTGGTAATCC 120

Qy    748 gtgcgcagcggtttactcctagcgaccatctggccaagtcgcgaggtgcgaaggaagct 807
          || | ||| |||| |||| || ||| ||||| ||||| |||| |
Db    121 TTGCGGGAAATGTCTATAGCTACCAGCATTTGATGAAGATTGCGGAATCGGCGAAAAAACT 180

Qy    808 cggaatattggtgatcgctgacgaggtttacggcaaactggttctcggcagcgcccgtt 867
          || || |||| ||| || |||||| || ||||| ||||| ||| |||||
Db    181 AGGGTTTCTTGTGAATTGGTGTGTTTGGATCTATTGNCCCTGCCTTTACTCTTGGCTCTTTATCAAA 240

Qy    868 tatcccgatgggcgtctttggcacattgccccggtcttgtccattggatctctgtccaa 927
          ||||| || |||| || || |||| |||||| ||||| || |
Db    241 TGTGCCAATGGGTGGTGTGTTGGATCTATTGNCCCTGCCTTACTCTTGGCTCTTTATCAAA 300

Qy    928 gtcgtggatagtgcctggatggcgactggatgtggtggcggtgtacgaccccac-aaaga 986
          |||| || || |||| |||||| | ||| | || ||| |||
Db    301 GAGATGGATAGTTCCAGGTTGGCGACTCGGGTGGTTTGTCACCACTGATCTTCTTGGTTC 360

Qy    987 ttttagagaaactaagatctctacgtctattacgattaccttaatgtctcaacggacc 1046
          ||| |||| || || |||| |||| |||| ||| ||
Db    361 CTTTAAAGGACCCTAAGATCATTGAGAGGTTTAAGAGAAATACTTGATAATTCTTGGTGGAC 420

Qy   1047 cagcaacctcgttcaggaagctctt 1072
          ||| || |||| ||||| || |
Db    421 CAGCTACATTTATTCAGGCTCGCAGTT 446
```

RESULT 7
T76370

```
LOCUS       T76370        470 bp    mRNA            EST        09-JAN-1998
DEFINITION  11148 Lambda-PRL2 Arabidopsis thaliana cDNA clone 149B19T7, mRNA
            sequence.
ACCESSION   T76370
NID         g935370
VERSION     T76370.1  GI:935370
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
            McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
            Retzel,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Apr 14, 1993 this sequence version replaced gi:693132.

            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7 dye primer.
FEATURES            Location/Qualifiers
     source         1..470
                    /organism="Arabidopsis thaliana"
                    /strain="var columbia"
                    /db_xref="taxon:3702"
                    /map="17 cen-qter"
                    /clone="149B19T7"
                    /clone_lib="Lambda-PRL2"
                    /note"Vector: lambda Zip-Lox: Site_1: Sal; Site_2: Not;
                    Lambda PRL2 is a cDNA library derived from equal
                    quantities of 4 pools of mRNA. The mRNA sources were 1) 7
                    day germinated etiolated seedlings; 2) tissue culture
                    grown roots; 3) staged plants half with 24 hour light
                    cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
                    same plants as 3 but aerial tissue (stems, flowers and
                    siliques. The vector is BRL's lambda Zip-Lox. The cDNA
                    inserts were directionally cloned with Sal-Not arms using
                    oligo dT primed cDNA."
BASE COUNT      116 a    104 c    110 g    122 t     18 others
ORIGIN
```

Query Match   8.2%; Score 136.8; DB 23; Length 470;
Best Local Similarity 65.4%; Pred. No. 1.3e-14;
Matches 212; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy    565 cgccacactgcttccccgccaggctatccaaattacgaggcggcgagcggcattcaa 624
          ||||| || || ||| || || || ||| |||||||| | || | |
Db      7 CGCAAACATCTTGCTCCCACGTCCTGGCTTCCCTCATTACGACGCTCGTNCNGCTTACAG 66

Qy    625 caagctggagtccggcacttcgacctcatcccgacaaggggtgggagatcgacatcga 684
          || || || ||| ||| || || |||||||| || |||||||| |||| | |
Db     67 TGGTCTCGAGGTTCGCAAGTTNATCTTCTTCCGGAAAAGAAATGGGAGATTGATCTTNA 126

Qy    685 ctcgcttggaatccatcgccgacaagaacaccacccgcgatggtcatcataaacccaaacaa 744
          || || ||| ||| || | |||||| || | |||| ||| ||| |||| ||| || |
Db    127 AGTATCGAAGCCATTGCAGACGAGACGAGAAAACTGTGGCTATGGTTGTAATTAACCCCAACAA 186

Qy    745 tccgtgcggcagcgtttactcctccagtaccatctggccaaggtcgcgaggtgcgcaaggaa 804
          |||| | || ||| || ||||| |||||||| ||||| ||||| || || || ||||||
Db    187 TCCCTGTGGAAATGTCTACTCTCACGACCATCTCAAAAAGGTTGCAGAGACCGGCTAGGAA 246

Qy    805 gctcggaatattggtgatcgctgacgaggt-ttacggcaaactggttctggcggcagcgccc 863
          ||||| |||| | |||||| ||||| || || || || |||||| ||| | ||| |
Db    247 GCTCGGGATAAATGGTGATCTCAGACGAAGTAATTATGACCGAACTATATTCGGAGACAATC 306

Qy    864 cgtttatcccgatgggcgtctttg 887
          ||||| ||| ||||| || |||
Db    307 CATTTGTTCCAATGGGGAAGTTTG 330
```

RESULT 8
D23043

```
LOCUS       D23043        343 bp    mRNA                        20-JUL-1998
DEFINITION  RICC2086A Rice callus Oryza sativa cDNA clone R, mRNA sequence.
ACCESSION   D23043
NID         g426967
VERSION     D23043.1  GI:426967
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 343)
  AUTHORS   Sasaki,T. and Minobe,Y.
  TITLE     Rice cDNA from callus
  JOURNAL   Unpublished (1994)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
```

Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.

FEATURES
source
1. .343
/organism="Oryza sativa"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2: NotI; cDNA prepared from rice callus mRNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site of pBluescript II SK+ phagemid."

BASE COUNT   90 a   58 c   80 g   94 t   21 others
ORIGIN

Query Match   7.4%;   Score 122.8;   DB 20;   Length 343;
Best Local Similarity   70.9%;   Pred. No. 3.2e-12;
Matches 207; Conservative   0; Mismatches 81; Indels   4; Gaps   4;

Qy   780 ccaaggtcgcggaggtggcaaggaagctcggaatattggtgatcgctgacgaggtttacg 839
          |||||| ||||||| |||||| |||||| || |||| || ||||| | |||||||||
Db   3 CCAAGGTGGCAGAGGTAGCAAGGAAGCTTGGGATATTGGTAATTACTGATGAGGTGTATG 62

Qy   840 gcaaactggttctgggcagcgcgccgtttatccgatgggcgtctttgggcacattgccc 899
          ||||| || |||||| |||| || ||||||||| || |||||||||| |||| |||||
Db   63 GTANTTTGGTTTTTGGGAGTTCCNCANTTGTCCCAANTGGTTACTTTGGGCACATCGTAC 122

Qy   900 cggtcttgtccattggatctctgtccaagtcgtggatagtgcc-tggatggcggacttg-g 957
          |||||||||||||||| || |||||| ||||| || ||||||| || |||| ||||||||
Db   123 CATAANNANCCNTAGGATCGCTATCAAAGAGGTGGATAGTGCCGTAGTGGCGACTTGAG 182

Qy   958 atgggtggcggtgtacgacccacaaagattttagagagaaaactaagatctc-tacgtccta 1016
          ||||||| ||||||| ||||||||||| ||||| | |||| |||| | |||||||||
Db   183 NTGGNTAGCANTATGTGACCCCAAGAGACTCTACAAGAGAAACCAAGATTGCANCCATTAA 242

Qy   1017 ttacgaattacctaatgtctcaac-gacccagcaacccttcgttcaggaag 1067
           |||||| ||||| |||||||| ||| ||||||| | ||||||||| ||
Db   243 TTACTAATTTCCTTAAATGGTTCAACTGNATCCAGCANCTTTCATTCAGGGAG 294

RESULT   9
T13684
LOCUS   T13684   352 bp   mRNA   EST   07-JAN-1998
DEFINITION 1849 Lambda-PRL2 Arabidopsis thaliana cDNA clone 35D4T7, mRNA sequence.
ACCESSION   T13684
NID   g930780
VERSION   T13684.1   GI:930780
KEYWORDS   EST.
SOURCE   thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 352)
AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel,E. and Somerville,C.
TITLE   Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
MEDLINE   95148729
COMMENT   On Nov 29, 1993 this sequence version replaced gi:430053.

Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
Lansing,Mi
Tel: 517-353-0854

Fax: 517-353-9168
Email: 22313tcn@ibm.cl.msu.edu
Seq primer: T7.

FEATURES
source
1. .352
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/map="4"
/clone="35D4T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site_1: SalI; Site_2: NotI; Lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark- rosettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques. The vector is BRL's lambda Zip-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA."

BASE COUNT   77 a   61 c   83 g   118 t   13 others
ORIGIN

Query Match   7.1%;   Score 117.8;   DB 23;   Length 352;
Best Local Similarity   63.1%;   Pred. No. 2.3e-11;
Matches 173; Conservative   0; Mismatches 101; Indels   0; Gaps   0;

Qy   690 tggaatcatcgcgacaagaacaccacccgatggtcatcataaaccaaacaatccgt 749
          ||| ||| | || ||| ||| ||| | || |||| || ||| ||| |||||
Db   2 TCGAGGCTCTTGCAGACACAAAACACGGTTGCTTTGGTTGTTATAAACCNGGTAATCCTT 61

Qy   750 gggcagcgttactcctacgaccatctggccaggtcgcgaggtggcaaggaagctcg 809
          ||| ||| ||||| |||| || || ||| || || || ||||| ||| ||||
Db   62 GCGGGAATNTCATAGCTACCAGCATTGATGAAGATTNCGGAATCGGCGAAAAAACTAG 121

Qy   810 gaatattggtgatcgctgacgaggtttacggcaaactgttctggcagcgcgccgttta 869
          ||| ||||| || ||| ||| ||||||| || |||| |||||||| |||| || ||
Db   122 GGTTCTTGTGATTNCTGATGAGGTTACGGTCATCTGCTTTGGTAGCAAACCGTTTG 181

Qy   870 tccgatgggcgtcttgggcacattgcccggtcttgtccattggatctctgtgtccaagt 929
          | || ||||| ||| ||||| ||| |||||||||||||||||| || ||| |||||
Db   182 TNCCAATNGGTGTGTTTGGATCATTGNCCTGCTTACTCTGGNTCTTTATCAAAGA 241

Qy   930 cgtggatagtgccgctgagatggcgacttggatgggt 963
          || |||| || || ||||| ||||| ||| |||| ||
Db   242 GATGGATAGTTCCAGGTTGCGACTCGGGTTNGT 275

RESULT   10
AA980369
LOCUS   AA980369   719 bp   mRNA   EST   27-MAY-1998
DEFINITION ua52c12.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone IMAGE:1350358 5' similar to gb:X52520_cds1 TYROSINE AMINOTRANSFERASE (HUMAN);, mRNA sequence.
ACCESSION   AA980369
NID   g3158905
VERSION   AA980369.1   GI:3158905
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 719)
AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE   The WashU-HHMI Mouse EST Project
JOURNAL   Unpublished (1996)
COMMENT   On Jan 19, 1998 this sequence version replaced gi:2152585.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:699150
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 463.

```
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Mus musculus"
                     /strain="C57BL/6 x CBA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1350358"
                     /clone_lib="Stratagene mouse lung 937302"
                     /sex="female"
                     /tissue_type="lung"
                     /dev_stage="6-8 month old"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: lung; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 6-8 month old female lung and 1.5 year old male
                     lung were source of mRNA. Average insert size: 1.5 kb;
                     Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACGAG
                     3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      160 a    185 c    195 g    178 t      1 others
ORIGIN
```

Query Match   6.4%;   Score 106.6; DB 40;   Length 719;
Best Local Similarity   53.2%;   Pred. No. 1.9e-09;
Matches 271; Conservative   0; Mismatches 234; Indels   4; Gaps   2;

```
Qy    488 gcgcgacgctcttcctcaccgccggactcaggcgatcgaagtcataatcccggtg 547
              || || |||| ||  ||| ||||| || |||||| | || ||||| || |||
Db      5 GCTAAGGATGTCATTCTGACGAGTGCTCGACGTCAGGCCATTGAGCTGTGTCTAGCCGTG 64

Qy    548 ctggcccagactgccggcgcccaacatactgcttccccggccagcctatccaaattacgag 607
              | |||||| || ||| ||| ||  ||||| ||  |||||| |||| | ||| || |||
Db     65 TTGGCC---AATCCTGGACAGAGAACATCCTCATTCCGAGGCCCGGTTTCCCTCTACAGG 121

Qy    608 gcgcgagcggcattcaacaagctgagggtccggcacttcgacctcatcccgacaaggggg 667
              | ||| | ||  ||| |||||||  ||||| |||||| | || |||  | ||| | ||
Db    122 ACATTGGCTGAGTCTGAGTGGGATTGAGGTCAAGCTCTACAATCATTGCCTGAGAAGTCT 181

Qy    668 tgggagatcgacatcgactcgtgaatccatcgccgacaagacaccacccgcgatggtc 727
              ||||  |||  | |||| |||  | |||| || ||| || || ||  | |||| ||||
Db    182 TGGGAAATTGATCTAAAACAACTGAATCTCTGATCGACGAGAAAAAACAGCTTGTCTCGTG 241

Qy    728 atcataaacccaaacaatccgtgcggcagcgttactcctacgaccatctggccaaggtc 787
              || | ||  ||| ||  || || || ||  ||| ||||| | |||||||| |||||| |
Db    242 GTCAACAACCCGTCCAATCCGTGTGCGCTCGTGTGTTCAGTAAGCGGCACCTTCAGAAGATT 301

Qy    788 gcggaggtggcaaggaagctcggaatattggtgatcgctgctacgagggttacggcaaactg 847
              || ||| || ||||| ||| | |  | ||||| | ||| || ||| |||  ||||| |||
Db    302 TTGGCAGTGGCTGAAAGGCAATGCGTCCCCATCTAGCCGATGAGATCATGGTGACATG 361

Qy    848 gttctgggcagcgccgcccgtttatcccgatggcgtctttggcacattgcccggtcttg 907
              | ||| |||||||  | ||| |||||| |||  ||||| | |  ||| | |||||| |
Db    362 GTGTTTTCAGATTGCAAAATATGAACCAATGGCCACCACCCTCAGCACCACCCATCCTG 421

Qy    908 tccattggatctctgtccagtcgtggatagtgcctggatgtgcgacttggcgactggtggcc 967
              |||| | || ||  ||| ||||| ||  |  |||| || || |||  | ||| ||||  |
Db    422 TCCTGTGGTGGGCTGGC-AAGCGCTGGTCTGTTCCTGGCTGGAGGCTGGGCTGGGATCCTT 480

Qy    968 gtgtacgaccccacaaagatttttagagaa 996
              ||||| ||| ||| ||| |||| | || |||
Db    481 ATCCATGATCGAAGAGACAATTTTTGGCAA 509
```

RESULT 11
AI621525/c
LOCUS   AI621525   283 bp   mRNA   EST   21-APR-1999
DEFINITION 486092E01.x1 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI621525
NID   g4630651
VERSION   AI621525.1 GI:4630651
KEYWORDS   EST.
SOURCE   Zea mays.
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1 (bases 1 to 283)
  AUTHORS   Walbot,V.
  TITLE   Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
  COMMENT   On Mar 20, 1998 this sequence version replaced gi:2980517.

            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486092 row: E column: 01.

```
FEATURES             Location/Qualifiers
     source          1..283
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /map="12; 4p16.1-4pter"
                     /clone_lib="486 - leaf primordia cDNA library from Hake
                     lab"
                     /tissue_type="leaf primordia"
                     /dev_stage="P7-P11 leaf"
                     /lab_host="E.coli XL1-Blue MFR'"
                     /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
                     library."
BASE COUNT       88 a     50 c     50 g     95 t
ORIGIN
```

Query Match   6.2%;   Score 102.4; DB 49;   Length 283;
Best Local Similarity   77.5%;   Pred. No. 1e-08;
Matches 124; Conservative   0; Mismatches 35; Indels   0; Gaps   0;

```
Qy   1293 aatcagtaatttatctccaggagtgttcttggaatggaaaattggtccgtattactt 1352
              |||| | || ||||||| |  ||||||||||||||||||||||||| |||||||||
Db    283 AGTCCGTGATTTTGTGTCCAGGAGTGTTTGGAATGGAAAACTGGATCCGTATCACTT 224

Qy   1353 ttgcctgcttccatctctcttcaagatggactcgaagggtcaaatcattcgtcaa 1412
              |||| |||||| ||| ||| ||||||| ||||| |||||| |||||| |||| |||
Db    223 TCGCCATTGATTCATCTCTCTTCTTGATGGTCTTGAGAGGCGAAATCTTTCTGCCAAA 164

Qy   1413 ggaacaagaagaattctataaatggttgttagttgta 1452
              ||| ||| |||| ||| ||||||||  || ||||| |||
Db    163 GGCATAAGAAGAAGAATTGCTTAATGGCCATTAACTATA 124
```

RESULT 12
AA511748
LOCUS   AA511748   512 bp   mRNA   EST   08-JUL-1997
DEFINITION vj29l11.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone IMAGE:930477 5' similar to gb:X52520_cds1 TYROSINE
            AMINOTRANSFERASE (HUMAN);, mRNA sequence.
ACCESSION   AA511748
NID   g2249602

```
Db  301  TTGGCAGTGGCTGAAAGGCAATGCGTCCCCATCTTAGCCGATGAGATCTATGGTGACATG  360
                ||||  || ||||||||  ||||||  ||| |||  ||| |||||||||||||||
Qy  848  gttctgggcagcgcccgtttatccgatgggcgtcttgggcacattgcccggtcttg  907
              ||  ||  |||  ||  ||||||  || |||| |||  ||||||  |||  ||
Db  361  GTGTTTCAGATTGCAAATATGAACCAATGGCCACCCCCAGCACCAATGTCCCATCCTG  420

Qy  908  tccattggatctcgtccaagtcgtggatagtcgtgatgtggcgacttggatg  961
              |||  |||  ||||  ||||||  ||||  || ||  ||  |||||||||
Db  421  TCCTGTGGTGGGGCTG-GCAAGCGCCTGGCTGGTTTCTCGGGCTGAGGCTGGGCTGG  473
```

RESULT 13
```
Z38047/c
LOCUS        Z38047        405 bp        mRNA        EST        12-OCT-1994
DEFINITION   ATTS4207 Versailles-VB Arabidopsis thaliana cDNA clone VBVMA10 3',
             mRNA sequence.
ACCESSION    Z38047
NID          g557616
VERSION      Z38047.1  GI:557616
KEYWORDS     EST.
SOURCE       thale cress.
  ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
             eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
             Arabidopsis.
REFERENCE    1  (bases 1 to 405)
  AUTHORS    CNRS.
  TITLE      The Arabidopsis thaliana transcribed genome:  the GDR cDNA program
  JOURNAL    Unpublished (1996)
COMMENT      Contact: Desprez T., Amselem J., Chiapello H., Rouze P., Caboche
             M., Hofte H.
             INRA Versailles
             Laboratoire de Biologie Cellulaire
             Route de Saint-Cyr,78026 Versailles Cedex,France
             Email: thierry@versailles.inra.fr.
FEATURES             Location/Qualifiers
     source          1..405
                     /organism="Arabidopsis thaliana"
                     /strain="ecotype Columbia"
                     /db_xref="taxon:3702"
                     /clone="VBVMA10"
                     /clone_lib="versailles-VB"
                     /note=Vector: pBluescript; tissue=whole seedlings of A.
                     thaliana ecotype Columbia; clone_library=Versailles-VB;
                     Cloning vector: pBluescript (Stratagene); Physiological
                     conditions: in vitro-grown etiolated seedlings,5 days
                     old."
BASE COUNT     132 a     80 c     73 g    117 t     3 others
ORIGIN

Query Match          5.5%;  Score 90.6;  DB 20;  Length 405;
Best Local Similarity  61.0%;  Pred. No. 1.1e-06;
Matches  147;  Conservative  0;  Mismatches  94;  Indels  0;  Gaps.  0;

Qy  1196  ccagaaggatcgatgttgtaatgtcaaactaaacttacatcttttggaggagatccat  1255
                |||| ||||||||||||||  |||| |||||| ||||| || ||||||||||||
Db   372  CCAGAAGGATCCATGGCAATGAGTGTTAAGCTGAATCTCTCATTACTTGAACATGTAAGT  313

Qy  1256  gacgacatagattttgctgcaagctcgcaaaggaagaatcagtaatttatgtcacagg  1315
                |||| ||  |||||||||||||| |||||||||||||||||||||||||||
Db   312  GACGATATCGACTTCGTTCAAGTTAGCTAGCTAGGGGAAGAATCAGTCAGTCACTTCTGGG  253

Qy  1316  agtgttcttgaatgaaaattgggccgtattactttggctggttccatcttcctt  1375
                ||| || |||| |||||||  ||| |||||||  |||||||| |||||||||
Db   252  ACCGGCGGTGGGGCTGAAGAACTGGCTGAGGATAACGTTGCAGGATGCAACTTCGATT  193

Qy  1376  caagatggactgaaggtcaaatcattcgtcaaggaacaagaagaattcatat  1435
                ||  ||||||||||||||| |||||||  | ||||| |||| ||||  |||
Db   192  GAAGAAGCTTTTAAAAGGATCAAATGTTTCATCTTAGACATGCCAAGACCTCAATATCCA  133
```

(Left column — RESULT 12 / earlier entry)

```
VERSION      AA511748.1  GI:2249602
KEYWORDS     EST.
SOURCE       house mouse.
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
             Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 512)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., and
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished (1996)
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1398101.
             Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of MedicineP
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:535397
             Seq primer: -28m13 rev1 ET from Amersham
             High quality sequence stop: 366.
FEATURES             Location/Qualifiers
     source          1..512
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:930477"
                     /clone_lib="Stratagene mouse diaphragm (#937303)"
                     /tissue_type="diaphragm"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
                     prepared from diaphragm muscle. Primer: Oligo dT. Average
                     insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTT 3'."
BASE COUNT     118 a    122 c    139 g    133 t
ORIGIN

Query Match          5.8%;  Score 95.6;  DB 34;  Length 512;
Best Local Similarity  53.0%;  Pred. No. 1.5e-07;
Matches  251;  Conservative  0;  Mismatches  219;  Indels  4;  Gaps  2;

Qy  488  gccgacgacgtcttcctcaccgccgcgggaactcaggcgatgaagtcataatcccggtg  547
              || ||  |  || ||| |||||||||||||  ||||||| |  |||||||||||||
Db    4  GCTAAGGATGTCATTCTGACAGTGGCTGCGTCAGCCATTGAGCTGTGTCTAGCCGTG  63

Qy  548  ctgcccagactgccggcgccaacatactgcttccccggccaggctatccaaattacgag  607
              ||||||| |||||| ||| ||| | ||| | |||||||||||| |||| ||| |||
Db   64  TTGGCC---AATCCTGACACAGAACATCCTGACGGGCCCGGTTTCCCTCTCTACAGG  120

Qy  608  gcgcgacgcggcattcaacaagctggagtccggcacttcggacctcatccccgacaagggg  667
              || ||  |||||||||  || ||||||| ||||||  ||| |||||||||||||||||||
Db  121  ACATTGGCTGCGAGTCTATGGGGATTGAGGTCAAGCTCTACAATCTATTGCCTGAGAAGTCT  180

Qy  668  tgggagatgcgacatcgagctcgctggagatccatcgcgcgacaagaacaccaccgcgatggtc  727
              |||||| ||| || || || |||  |||| ||||| | ||| |  |||
Db  181  TGGGAAATTGATCTAAAAACAACTGGAATCTGATCGACGAAAAAACAGCAGTTGTCTCGTG  240

Qy  728  atcataaacccaaacaatcgtgcgggcagcgttactcctacgaccatctggccaagtc  787
              ||  |||   ||||||||  || ||| |  ||||| || ||||||| |||||||||||
Db  241  GTCAACAACCCGTCCAATCCCGTCGTGGCTCGTGTTCAGTAAGCGGCACCTTCAGAGATT  300

Qy  788  gcggaggtggcaaggaagctcggaatatatggtgatcgctgacgaggttacgcaaactg  847
              |  |||||| || || ||||||||||  ||||||||||||||  || || ||| |||
Db  ...
```

```
Qy   1436 a 1436
          |
Db    132 A 132
```

RESULT 14
T41757

```
T41757
LOCUS       T41757       347 bp     mRNA                EST       07-AUG-1995
DEFINITION  10338 Lambda-PRL2 Arabidopsis thaliana cDNA clone 67A8T7, mRNA
            sequence.
ACCESSION   T41757
NID         g931391
VERSION     T41757.1   GI:931391
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H.,
            McIntosh,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M.,
            Retzel,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Jan 24, 1995 this sequence version replaced gi:634345.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,MI
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Insert Length: 357   Std Error: 0.00
            Seq primer: T7 dye primer
            High quality sequence stop: 250.
FEATURES             Location/Qualifiers
     source          1..347
                     /organism="Arabidopsis thaliana"
                     /strain="var Columbia"
                     /db_xref="taxon:3702"
                     /clone="67A8T7"
                     /clone_lib="Lambda-PRL2"
                     /note="Vector: lambda Zip-Lox; Site_1: Sal; Site_2: Not;
                     Lambda PRL2 is a cDNA library derived from equal
                     quantities of 4 pools of mRNA. The mRNA sources were 1) 7
                     day germinated etiolated seedlings; 2) tissue culture
                     grown roots; 3) staged plants half with 24 hour light
                     half on 16 hr light, 8 hour dark- rosettes; 4)
                     same plants as 3 but aerial tissue (stems, flowers and
                     siliques. The vector is BRL's lambda Zip-Lox. The cDNA
                     inserts were directionally cloned with Sal-Not arms using
                     oligo dT primed cDNA."
BASE COUNT      81 a     68 c     78 g    103 t     17 others
ORIGIN
```

Query Match 5.2%;   Best Local Similarity 55.2%;   Score 86.8;   Pred. No. 4.8e-06;   DB 23;   Length 347;
Matches 174;   Conservative 0;   Mismatches 140;   Indels 1;   Gaps 1;

```
Qy  710 aacaccaccgatggtcatcataaccaaacccaaaatccgtgcggcagcgttactcctac 769
         | |||  ||| ||||| |||||||  | |||| |  ||||| ||| |  ||| |||  |
Db    2 AAAACCGTTGCTATACTTGTNATCAACCCTTGCAATCCATGTGGGAATGTTTTNTCTCGC 61

Qy  770 gaccatctggccaagtcgcgagagtggcaaggaagctcggaagaattggtatgctgac 829
         ||| ||||| ||| || ||||| | || ||| ||||  |||| ||||| ||| ||||
Db   62 CAACATCTTCAAAAGATTGCAGAGACGCGCTTGCAAGCTTGGTATACTTGTNATCGGAGAC 121
```

```
Qy  830 gaggtttacggcaactggttctggcgcagcgcccgtttatcccgatgggggcgtctttggg 889
         | |||||| || ||||  ||| ||| | ||| || |  | | ||||||  | | ||||
Db  122 GAAGCTATGACCATTTTNCCTTGGGGA-TAAACCATTTGTNTCCATGGCAGAGAGTTTGCA 180

Qy  890 cacattgcccggtcttgtccattggatctgtccaagtcgtggtagtgcctggatgg 949
         ||  | ||| || ||||||| ||| |||||||||| ||||| |||| |||||||||
Db  181 GAGCTAGTNCCTGTNATTGTNTTAGGTGCTATATCTAAAAGATGGTTGTNCCNGGATGG 240

Qy  950 cgacttggatggtgcggttgtacgacccacaaagatttagagaaaactaagatcctct 1009
         | |||||| |||||| |  ||  |||| || |||| | ||||| | ||||  ||| |
Db  241 AGACTTGGTTGGATGGTGACNCTTGACCCTCATGGCATCATGAAAGATTCTGGGTTTNTN 300

Qy 1010 agtcttattacgaat 1024
         ||| ||||| ||||
Db  301 CAGAGCTCTTATCAAT 315
```

RESULT 15
AI597379

```
AI597379
LOCUS       AI597379     415 bp     mRNA                EST       21-APR-1999
DEFINITION  vj29f11.y1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone IMAGE:930477 5' similar to gb:X52520_cds1 TYROSINE
            AMINOTRANSFERASE (HUMAN);, mRNA sequence.
ACCESSION   AI597379
NID         g4606427
VERSION     AI597379.1   GI:4606427
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 415)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson.R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     On Jun 5, 1998 this sequence version replaced gi:3189286.
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:535397
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 399.
FEATURES             Location/Qualifiers
     source          1..415
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:930477"
                     /clone_lib="Stratagene mouse diaphragm (#937303)"
                     /tissue_type="diaphragm"
                     /dev_stage="adult"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
                     prepared from diaphragm muscle. Primer: Oligo dT. Average
                     insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
                     sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTT 3'."
BASE COUNT     104 a    102 c    103 g    106 t
ORIGIN
```

us-09-026-400-1.rst

Fri Oct 1 16:47:11 1999

```
Query Match      5.1%;  Score 85;  DB 48;  Length 415;
Best Local Similarity  52.1%;  Pred. No. 9.7e-06;
Matches   215;  Conservative   0;  Mismatches 195;  Indels   3;  Gaps   1;

Qy  488 gccgacgacgtcttcctcaccgccggcggaactcaggcgatcgaagtcataatcccggtg 547
            |!  | | | | | | | |!  ! |   | |  | | | | | ! !! ! ! !  ! !!!
Db    4 GCTAAGGATGTCATTCTGACGAGTGGCTGCAGTCAGGCCATTGAGCTGTGTCTAGCCGTG  63

Qy  548 ctggcccagactgccggcgccaacatactgcttccccggccaggctatccaaattacgag 607
            | | | |         | | | | !    | | | |   | |  | | |  | |  |  |
Db   64 TTGGGC--AATCCTGGACAGAACATCCTCATTCCGAGGCCCGGGTTTTCCCTCTACAGG 120

Qy  608 gcgcgagcggcattcaacaagctggaggtccggcacttcgacctcatcccgacaagggg 667
            | | |  | |           | |  || |!  | | | | | | | |! || | |!!
Db  121 ACATTGGCTGAGTCTATGGGGATTGAGGTCAAGCTCTACAATCTATTGCCTGAGAAGTCT 180

Qy  668 tgggagatcgacatcgactcgctgtggaatccatcgccgacaagaacaccaccgcgatggtc 727
            | | | | | |  |  | | |      !! | | | | | |      !  ! |  |  | ||
Db  181 TGGGAAATTGAICTAAAACAACTGGAATCTCTGATCGACGAAAAACAGCTTGTCTCGTG 240

Qy  728 atcataaacccaaacaatccgtgcggcagcgtttactcctacgaccatcctggccaaggtc 787
            | | | | | | | | |   | | | | | |      | |        | | |    | | | |
Db  241 GTCAACAACCCGTCCAATCCCTGTGCTCGTGGTTCAGTAAGCGGCCACCTTCAGEAAGATT 300

Qy  788 gcggaggtggcaaggaagtcggaatattggtgatcgctgacgaggtttacggcaaactg 847
            |  | | | | | | |    |  |  | | | | | | | |  |  | | | |  | | | | |
Db  301 TTGGCAGTGGCTGAAAGGCAATGCGTCCCCATCTTAGCCGATGAGATCTATGGTGACATG 360

Qy  848 gttctggcagcgccccgtttatcccgatgggcgtctttgggcacattgcccc 900
            |  |  |    | | |  | |       | | | | | | |   | | | | | | !
Db  361 GTGTTTTCAGATTGCAAATATGAACCAATGGCCACCCTCAGCACCAATGTCCC 413
```

Search completed: October 1, 1999, 15:03:30
Job time: 4374 sec

GenCore version 4.5
Copyright (C) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 1999, 13:51:37 ; Search time 1619.02 Seconds
(without alignments)
3260.808 Million cell updates/sec

Title: US-09-026-400-1
Perfect score: 1660
Sequence: 1 attgactagctagttcattc........aaaaaaaaaaaaaaaa 1660

Scoring table: IDENTITY_NUC
Searched: 679419 seqs, 159015680 residues

Database : GenEmbl:*
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_prl:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_bal:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	144.2	8.7	2362	12	RNTATR	X02741 Rat mRNA fo

No.	flag	Score	%	Length	N	ID	Description
2		137.4	8.3	103495	8	ATAC006585	AC006585 Arabidops
3		134.6	8.1	2004	12	RATTATRA	M18340 Rat tyrosin
4		133.6	8.0	93695	7	ATF2009	AL021749 Arabidops
5		126.2	7.6	61384	8	ATAC007048	AC007048 Arabidops
6		125	7.5	2754	10	HSTATR	X52520 Human mRNA
7		125	7.5	2051	10	HSTYAT	X55675 H.sapiens m
8		118.2	7.1	73840	7	AB007644	AB007644 Arabidops
9	c	110.2	6.6	39524	9	SCH10	AL049754 Streptomy
10	c	110.2	6.6	70475	7	AB018112	AB018112 Arabidops
11	c	107.2	6.5	25970	1	SC2H4	AL031514 Streptomy
12	c	104.8	6.3	3240	1	SAAJ5198	AJ005198 Streptomy
13	c	104.8	6.3	3240	1	SAN6517	AJ006517 Streptomy
14	c	95	5.7	39399	1	MSGB1970CS	L78815 Mycobacteri
15	c	94.2	5.7	119430	7	ATF9D16	AL035394 Arabidops
16	c	93	5.6	9150	1	MTCY279	Z97991 Mycobacteri
17	c	89.2	5.4	7409	1	STMVBRA1	D50624 Streptomyce
18	c	88.6	5.3	32258	36	CEF42D1	Z81081 Caenorhabdi
19	c	87.8	5.3	187647	34	CEY16B4	AL021148 Caenorhab
20		73.6	4.4	1663	1	TAX99521	X99521 T.aquaticus
21		67.2	4.0	1777	1	TTHASPC	D38459 Thermus the
22	c	67	4.0	2977	1	AF025665	AF025665 Thermus a
23		67	4.0	7235	1	SGNUSG	X72787 S.griseus n
24		65.6	3.9	5732	1	STMMUTA	L10064 Streptomyce
25	c	65.4	3.9	3824	9	HUMOCTF1A	L20433 Human octam
26	c	64	3.9	1149	9	HSBRN3A2	U10063 Human POU d
27	c	63.6	3.8	38532	1	SCE94	AL049628 Streptomy
28	c	63.6	3.8	3492	10	HSRDC1MR	X64624 H.sapiens m
29	c	63.6	3.8	3492	14	G28612	G28612 human STS s
30	c	63.4	3.8	22449	1	SPSNBCDE	Y11548 S.pristinae
31		63.4	3.8	22449	1	SPSNBCGEN	X98690 S.pristinae
32	c	63.4	3.8	2028	8	SSI132828	AJ132828 Spermatoz
33		62.8	3.8	17533	1	D90860	D90860 E.coli geno
34		62.8	3.8	10863	1	AE000318	AE000318 Escherich
35	c	62	3.8	37445	1	SC8A6	AL031013 Streptomy
36		62	3.7	53784	2	AMM223012	AJ223012 Amycolato
37		62	3.7	76199	2	AF040570	AF040570 Amycolato
38		62	3.7	53789	5	A69720	A69720 Sequence 3
39		61.6	3.7	2767	1	SSI224970	AJ224970 Streptomy
40	c	61.4	3.7	22789	1	SC2A11	AL031184 Streptomy
41		61	3.7	67200	1	MTV017	AL021897 Mycobacte
42		61	3.7	3400	3	CSCLOXA	X69017 Canis sp. m
43		60.2	3.6	2984	36	TRBANTA	L00673 Trypanosoma
44		59.6	3.6	1166	1	STMNUSG	D17465 S.coelicolo
45		59.6	3.6	1506	4	CHKMARKS	M31650 Chicken myr

ALIGNMENTS

RESULT   1
RNTATR
LOCUS   RNTATR   2362 bp   mRNA   ROD   12-SEP-1993
DEFINITION Rat mRNA for tyrosine aminotransferase (L-tyrosine: 2-oxoglutarate
            aminotransferase EC 2.6.1.5).
ACCESSION   X02741 X12522 X15375
NID   g57327
VERSION   X02741.1 GI:57327
KEYWORDS   transferase; tyrosine aminotransferase.
SOURCE   Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Rodentia; Sciurognathi; Murinae; Rattus.
REFERENCE   1 (bases 1 to 2362)
AUTHORS   Grange,T., Guenet,C., Dietrich,J.B., Chasserot,S., Fromont,M.,
            Befort,N., Jami,J., Beck,G. and Pictet,R.
TITLE   Complete complementary DNA of rat tyrosine aminotransferase
            messenger RNA. Deduction of the primary structure of the enzyme
JOURNAL   J. Mol. Biol. 184 (2), 347-350 (1985)
MEDLINE   85293103
REFERENCE   2 (bases 1 to 2362)
AUTHORS   Dietrich,J.B., Genot,G. and Beck,G.
TITLE   Structural and immunochemical properties of rat liver tyrosine
            aminotransferase

```
Db    754  TCCCTGGCTCCGTGTTCAGTAAGCGACACCTTCAGAAGATTTGGCAGTGGCTGAAAG  813
            |  |||  ||| ||| ||  ||||| |  ||||||  ||||  ||||| ||||||||
Qy    805  gctcggaatattggtgatcgctgacgaggttacggcaaactggttctgggcagcgcccc  864
            |  |  |||||||||||    ||||| |  ||||||||  |  | ||| |  |||
Db    814  GCAGTGTGTCCCATCTAGCTGACGAGATCTATGGTGACCATGGTGTTTCAGATTGCAA  873
            | |||| ||||| ||||| |||||  ||||||||| |||  |  |  | |||||||
Qy    865  gtttatcccgatgggcgtcttggcacattgccccggtctgtcattgatcctgtc  924
            ||||  ||||||| |  |  ||||| | |  ||||||| ||  | ||| ||
Db    874  ATACGAACCACTGGCCAACCTCAGCACCAATGTTCCCATCCTGTCCTGTGGCTGGGC  933
            |  |||||| ||||||||  |||  | |||| |||  |||| |||  ||||||
Qy    925  caagtcgtggatagtgccgacttggatggtgcggtgctacgacccacaa  984
            ||  | | ||| || ||||||||||| || | |  |  ||  ||||  |||
Db    934  CAAGCCGCTGGCTGGTTCCTGGCTGGAGGTTGGGCTGCATCCTCATTCATGATCGAAGAGA  993
            ||||  ||||||| ||| ||| |||||||||||| |||  |||| |||| |||||||||
Qy    985  gatttagagaaaactaagatcctcacgtctattacgaattacctaatgtctcaacgga  1044
            ||||| ||||||||||    |||| | |||| | |  ||||  | ||||| | ||||
Db    994  CATTTTTGGCAA---TGAGATTCGAGACGGCTGGTGAAACTGACTGTCAGCGGATCCTCGG  1050
            |||||| |||||    ||| |||||||| ||||||| |||| |  ||||||  ||||||
Qy   1045  cccagcaacctcgttcaggagctcttcctaaaattcttgagaacacaaaagcagattt  1104
            |||| ||||  ||| |  |||||| |||  | |||| | ||||| |||| | ||||
Db   1051  ACCATGACCACCAGTCCAGGGTGCTCTGAAGAGCATCCTTCAGCGAGACCCCTCCAGGAGTT  1110
            ||||||| |||| ||||||| || ||  || || ||| |||||||  |||  | |||||
Qy   1105  ctttaagaggattattggtctactaaggaatcatcagagatatgttataggaaataaa  1164
            ||| |  |  |  ||  |  |  ||||| |  || ||  || |  ||| ||||||
Db   1111  CTATCACGACACGTTAAGCTTCCTCCAGTCCAATGCAGGACCTCGCTATGGGCACT---  1168
            ||  | ||| |  ||| |||||  |||||  | ||||||  ||   |||||||
Qy   1165  ggaaacaaatatattacgtgtcctcacaagccagaaggatcgatgtttgtaatggtcaa  1224
            ||||| ||| ||||| ||||| |||||| | | ||||||||||| ||| ||| | |||
Db   1168  GGCTACCCATCCCTGGACTCCAGCCGGTGCGCCCTTCGGAGCCATGCATGTTACCTTATGGTGGG  1227
            ||||  ||||  ||||  ||||| ||||| |  |||  |||||  ||||| || ||||||
Qy   1225  actaaacttacatcttttggaggagatccatgacgacgacatagatt  1269
            ||| ||| || ||||||||||||||| ||  ||||||||| |||||||
Db   1228  AATTGAGATGGAGCATTTCCCGGAATTCGAGAACGACGTGGAGTT  1272
```

```
RESULT 2
ATAC006585
LOCUS       ATAC006585   103495 bp   DNA         PLN    04-APR-1999
DEFINITION  Arabidopsis thaliana chromosome II BAC F27C12 genomic sequence,
            complete sequence.
ACCESSION   AC006585
NID         g4559344
VERSION     AC006585.6   GI:4559344
KEYWORDS    HTG.
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 103495)
AUTHORS     Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
            Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
            Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
            Fraser,C.M. and Venter,J.C.
TITLE       Arabidopsis thaliana chromosome II BAC F27C12 genomic sequence
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 103495)
AUTHORS     Lin,X. and Kaul,S.
TITLE       Direct Submission
JOURNAL     Submitted (19-FEB-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE   3  (bases 1 to 103495)
AUTHORS     Lin,X.
TITLE       Direct Submission
JOURNAL     Submitted (04-APR-1999) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Apr 4, 1999 this sequence version replaced gi:4522000.
            Address all correspondence to:
            Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr.
            Rockville, MD 20850, USA
```

```
JOURNAL    Biochimie 70 (5), 673-679 (1988)
MEDLINE    89000971
COMMENT    Data kindly reviewed (07-MAR-1986) by C. Guenet.
FEATURES   Location/Qualifiers
source     1..2362
           /organism="Rattus norvegicus"
           /db_xref="taxon:10116"
misc_RNA   1
           /note="cap-site"
CDS        98..1462
           /note="tyrosine aminotransferase (aa 1-454)"
           /codon_start=1
           /protein_id="CAA26519.1"
           /db_xref="GI:57328"
           /db_xref="PID:g57328"
           /db_xref="SWISS-PROT:P04694"
           /translation="MDSYVIQTDVDDSLSSVLDVHVNIGGRNSVQGRKKGRKARWDVR
           PSDMSNKTFNPIRAIVDNMKVQPNPNKTVISLSIGDPTVFGNLPTDPEVTQAMKDALD
           SGKYNGYAPSIGYLSSREEVASYYHCHEAPLEAKDVILTSGCSQAIELCLAVLANPGQ
           NILIPRPGFSLYRTLAESMGIEVKLYKLLPEKSWEIDLKQLESLIDEKTACLVVNNPS
           NPCGSVFSKRHLQKILAVAEROCVPILADEIYGDMVFSDCKYEPLANLSTNVPILSCG
           GLAKRWLVPGWRLGWLIHDRRDIFGNEIRDGLVKLSQRIIGPCTIVQGALKSIIQRT
           PQEFYHDTLSFLKSNADLCYGALAAIPGLQPVRPSGAMYLMVGIEMEHPFEFENDVEF
           TERLIAEQAVHCLPATCFEYPNFFRVVITVPEVMMLEACSRIQEFCEQHYHCAEGSQE
           ECDK"
misc_feature  2023..2028
           /note="weak polyA signal"
misc_feature  2345..2350
           /note="strong polyA signal"
polyA_site  2362
           /note="polyA site"
BASE COUNT   598 a   583 c   571 g   610 t
ORIGIN
```

```
Query Match        8.7%;   Score 144.2;   DB 12;   Length 2362;
Best Local Similarity  49.3%;   Pred. No. 1.8e-12;
Matches   495;   Conservative   0;   Mismatches   498;   Indels   12;   Gaps   4;

Qy   265  cagcgcgagcgtggaggagagcggcgcggccgctgctgctgctgcccacgtgaccc  324
           |  ||||| |   |||||||| |||||| | ||  | ||  ||||||  ||||||
Db   280  CAACATGAAGGTGCAGCCCAATCCGAACAAGACCGTGATTTCTCTGTCAATTGGGACCC  339
           |  |||| |||| | |  || || ||||||  |  ||| ||| || || | ||||
Qy   325  gtccgtgttcccggccttccgcacggccgtcgagcgccgaagacgccgtcgcgccgct  384
           | ||||| |||  ||  ||| || |  ||| ||  | ||  | | ||||||  |
Db   340  TACTGTGTTTGGGAACCTGCCTACAGACACCCTGAAGTTACCCAAGCCATGAAAGATGCCCT  399
           || ||| |||||||| ||  || | | ||| |||| |||||||  ||||  ||||||
Qy   385  gcgcacggccagttcaactgctacgcgcggccgtcggcctcccccggccgcgaagcgc  444
           || ||||| |  ||  | ||  ||||||  ||  |  ||  |  ||| ||||||||
Db   400  GGACTCGGGAAGTACAATGGCTATGCCCGTCCATGCCCTACCTATCCAGTCGGGAGGA  459
           |||  |||||| |  ||  ||||  ||||| ||  ||  || ||||  |||||||
Qy   445  cgtagcagagcacttgtcacagggcgtgccctacaagctatggccgacgagctcttct  504
           | ||||| |  | |||| | |||| || |  |  ||| | ||| ||||||  |||
Db   460  GGTCGCTTCTACTACCACTGCTCATGAGGCTCC---TCTGGAAGCTAAGGATGTCATTCT  516
           |||| || ||| ||  | |||||||| |||     |||||  |||| ||| ||||
Qy   505  caccgccggaactcaggcgatcgaagtcataatcccggtgctggcccagactgccgg  564
           || ||| ||||  |||| |||  || || || || | |||||| |||||  ||||||
Db   517  GACAAGCCGGCTGCAGTTCAGGCCATTGAGCTAGTCTGGTGTGTTGGCCAATCCT---GG  573
           ||  ||| ||| ||||  |||| ||||  || || | ||||| | |||||| |||
Qy   565  cgccaacatactgcttcccggccaggctatccaaattacggagccggaggcggcattcaa  624
           |||||| ||||| ||  || || |||| ||  ||  | | ||||||||| |||||||
Db   574  ACAAAACATCCCATTCCAAGGCCCGGTTTCCCTCTATAGAGCTTGGCTGAGTCTAT  633
           ||  || ||  |||  |||||| ||| |||  |||  | | ||| || |||||
Qy   625  caagctgaggtccggcacttcgacctcatccccgacaagggggagatcgacatcga  684
           ||||||| | |||||| ||| || | ||| || |||  ||| |||  ||| |||||
Db   634  GGGAATTGAGGTCAAGCTCTACAATCTCCTGCCTGAGAAGTCTTGGGAAATTGACCTAAA  693
           |||||| ||||| | ||| ||||| ||||  |||||||| || || |||||| ||
Qy   685  ctcgctggaatccatcgccgacaagaaccaccgcgatggtcatcataaacccaaacaa  744
           ||  |||||| ||  || || |  |  | ||| || || |||||  |  ||||  |||
Db   694  ACAACTGGAATCTCTGATCGATGAAAAAACAGCGTGCTTCGTTGTCAACAAACCCATCCAA  753
           ||  |||||| ||  ||||  || | |  || ||  |  |  ||||||| || |||
Qy   745  tccgtgcggcagcgtttactcctacgaccatctgcccacgaccatcttggagg  804
           | ||| || ||||  || | | ||| |||| |  | || ||||| | ||||
```

```
       e-mail: xlin@tigr.org
       BAC clone F27C12 is from Arabidopsis chromosome II and is near the
       molecular marker mi238.
       The orientation of the sequence is from SP6 to T7 end of the BAC
       clone.
       Genes were identified by a combination of three methods: Gene
       prediction programs including GRAIL (available by anonymous ftp
       from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
       Washington), Genscan (Chris Burge,
       http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
       (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
       complete sequence against a peptide database and the Arabidopsis
       EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
       Annotated genes are named to indicate the level of evidence for
       their annotation. Genes with similarity to other proteins are named
       after the database hits. Genes without significant peptide
       similarity but with EST similarity are named as 'unknown' proteins.
       Genes without protein or EST similarity, that are predicted by more
       than two gene prediction programs over most of their length are
       annotated as 'hypothetical' proteins. Genes encoding tRNAs are
       predicted by tRNAscan-SE (Sean Eddy,
       http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
       identified by repeatmasker (Arian Smit,
       http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
       genomic sequence that are not annotated as genes but have predicted
       exons by GRAIL are annotated as misc features.

FEATURES             Location/Qualifiers
     source          1..103495
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="mi238"
                     /clone="F27C12"
     repeat_region   100..136
                     /rpt_family="AT_rich"
     repeat_region   206..243
                     /rpt_family="AT_rich"
     misc_feature    complement(245..344)
                     /note="exon predicted by xgrail, quality
                     excellent_shadowexon"
     misc_feature    262..349
                     /note="exon predicted by xgrail, quality excellent"
     repeat_region   complement(353..418)
                     /rpt_family="AT_rich"
     mRNA            join(1069..1197,1624..1742,1826..1889,2061..2169,
                     2257..2410,2506..2755,2850..2926,3010..3062,3149..3203,
                     3287..3608)
                     /gene="F27C12.1"
     gene            1069..3608
                     /gene="F27C12.1"
     CDS             join(1626..1742,1826..1889,2061..2169,2257..2410,
                     2506..2755,2850..2926,3010..3062,3149..3203,3287..3475)
                     /gene="F27C12.1"
                     /codon_start=1
                     /product="protein phosphatase 2C"
                     /protein_id="AAD23006.1"
                     /db_xref="PID:g4559345"
                     /db_xref="GI:4559345"
                     /translation="MGTYLSSPKTEKLSEDGENDKLRFGLSSMQGWRATMEDAHAAIL
                     DLDDKTSFFGVYDGHGGKVVAKFCAKYLHQQVISNEAXKTGDVETSLRRAFFRMDDMM
                     QGQRGWREIAVLGDKMNKFSGMIEGFIWSPRSGDTNNQPDSWPLEDGPHSDFTGPTSG
                     CTACVALIKDKKLFVANAGDSRCVISRKSQAYNLSKDHKPDLEVEKERILKAGGFIHA
                     GRINGSLNLTRAIGDMEFKQNKFLPSEKQMVIADPDINTIDLCDDDFLVVACDGIWD
                     CMSSQELVDFIHEQLKSETKLSTVCEKVVDRCLAPDTATGEGCDNMTILVQFKKPNP
                     SETEPEDSKPEPSEDEPSSSS"
     misc_feature    3939..3988
                     /note="exon predicted by xgrail, quality
                     excellent_shadowexon"
     mRNA            complement(join(<4062..4414,4489..>4666))
                     /gene="F27C12.2"
     gene            complement(<4062..>4666)
                     /gene="F27C12.2"
     CDS             complement(join(4062..4414,4489..>4666))
                     /gene="F27C12.2"
                     /codon_start=1
                     /product="early nodulin 16"
                     /protein_id="AAD23007.1"
                     /db_xref="PID:g4559346"
                     /db_xref="GI:4559346"
                     /translation="MASSLHVAIFSLIFLFSLAAANEVTVGGKSGDWKIPPSSSYSF
                     TEWAQKARFKVGDFIVFRYYESGKDSVLEVTKFAYNSCNTTNPLANYTDGETKVKLDRS
                     GPFYFISGANGHCEKGQKLSLVVISPRHSVISPAPSPVFEFEDGPALAPAPISGSVRLG
                     GCYVVLGLVLGLCAWF"
     misc_feature    complement(5439..5548)
                     /note="exon predicted by xgrail, quality good"
     repeat_region   5777..5857
                     /rpt_family="AT_rich"
     misc_feature    complement(5974..6004)
                     /note="exon predicted by xgrail, quality excellent"
     repeat_region   complement(6006..6052)
                     /rpt_family="AT_rich"
     misc_feature    complement(6248..6315)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    complement(6332..6545)
                     /note="exon predicted by xgrail, quality good"
     misc_feature    6795..6873
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    7064..7159
                     /note="exon predicted by xgrail, quality marginal"
     mRNA            join(<7609..8283,8382..8603,8902..>10233)
     gene            <7609..>10233
                     /gene="F27C12.3"
     CDS             join(7609..8283,8382..8603,8902..10233)
                     /gene="F27C12.3"
                     /codon_start=1
                     /product="putative extensin"
                     /protein_id="AAD23008.1"
                     /db_xref="PID:g4559347"
                     /db_xref="GI:4559347"
                     /translation="MLEDEDYRVVYSRIMSQLREQFPGASFMVNFRDGDSRSRMESV
                     LTEYDMTIMDYPRHYEGCPLLTMETVHHFLKSAESWLLLSQQNILLSHCELGGWPTLA
                     FMLASLLLYRKQFSGEHRTLEMIYKQAPRELLQLMSPLNPLPSOLRFLQYISRRNVGS
                     QWPPLDQALILDCVNLRLIPDFDGEGGCRPIFRIYGQDPFMASDRTSKVLFSMPKRSK
                     AVRQYKQADCELVKIDINCHILGDVVLECITLGSDLEREMMFRVVFNTAFLRSNILT
                     LNRGEIDVLWNTTDRFPKDFSAEVIFSEMGAGKKLASVDLPHMEEKDVLPMEAFAKVQ
                     EIFSEAEWLDPNSDVAVTVENQIRAANIIQESLDSGSPRSPDSRSLLESALEKVKEKT
                     KLMISENIVSSPDTSSPEKERDTMSSHKKSYADPNSILKKVDESRGLRVSVQRNVHSKI
                     FSPRMVQSPVTSGPLPNRSPTQGSPASISRFHSSPSSLGITISILHDHGSCKDEESTSSS
                     PASPSISFLPTLHPLISSQPKKASPQCPQSPTPVHSNGPPSAEAAVTSSPLPPLKPLR
                     ILSRPPPPPPPPISSLRSTPSPSSTSNSIATQGPPPPPPPPLQSHRSALSSSPLPP
                     PLPPKKLLATNPPPPPPPPLHSNSRMGAPTSSLVLKSPPVPPPAPAPLSRSHNGNI
                     PPVPGPPLGLKGRGILQNLKGOGQTRKANLKPYHWLKLTRAVQGSLWAEAQKSDEAAT
                     YD"
     repeat_region   complement(9805..9935)
                     /rpt_family="(TGG)n"
     misc_feature    10313..10439
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    10569..10648
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    10804..10902
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    complement(10997..11102)
                     /note="exon predicted by xgrail, quality
                     excellent_shadowexon"
     misc_feature    11037..11112
                     /note="exon predicted by xgrail, quality good"
     misc_feature    11350..11404
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    11494..11575
                     /note="exon predicted by xgrail, quality good"
     misc_feature    11703..11797
                     /note="exon predicted by xgrail, quality excellent"
```

```
misc_feature   11966..12037
               /note="exon predicted by xgrail, quality excellent"
misc_feature   12135..12251
               /note="exon predicted by xgrail, quality excellent"
misc_feature   12341..12409
               /note="exon predicted by xgrail, quality excellent"
misc_feature   12524..12620
               /note="exon predicted by xgrail, quality excellent"
misc_feature   12742..12884
               /note="exon predicted by xgrail, quality excellent"
misc_feature   complement(12902..12986)
               /note="exon predicted by xgrail, quality excellent"
misc_feature   complement(14951..14971)
               /note="exon predicted by xgrail, quality
               marginal_shadowexon"
repeat_region  complement(join(<15100..15130,15200..15393,15508..15558,
               15975..>16006))
               /rpt_family="AT_rich"
               /gene="F27C12.4"
mRNA

CDS            98..1462
               /note="tyrosine aminotransferase (EC 2.6.1.5)"
               /codon_start=1
               /protein_id="AAA42203.1"
               /db_xref="PID:g207156"
               /db_xref="GI:207156"
               /translation="MDSYVIQTDVDDSLSSVLDVHVNIGGRNSVQGRKKGRKARWDVR
               PSDMSNKTFNPIRAKTDNMKVQPNPNKTVISLSIGDPFVFGNLPTDPEVTQAMKDALD
               SGKYNGYAPSIGYLSSREVASYHCHEAPLEAKDVILTSGCSQAIELCLAVLANPGQ
               NILIPRPGFSLYRTLAESMGIEVKLYNLLPEKSWEIDLKQLESLIDEKLECLVVNNPS
               NPCGSVFSKRHLQKILAVAERQCVPIIADEIYGDMVFSDCKYEPLANLSTNVPILSCG
               GLAKRWLVGGWRLGWILIHDRRDIFGNEIRDGLVKLSQRILGPCTIVOGALKSILQRT
               POEFYHDTLSFLKSNADLCYGALADLPGLQPVRPSGAMYLMVGIEMEHFPEFENDVEF
               TERLIAEQAVHCLPATCFEYPNFFRVVITVPEVMLEACSRIQEFCEQHYHCDEGSQE
               ECDK"
```

```
BASE COUNT     491 a   521 c   517 g   475 t
ORIGIN

Query Match        8.1%;   Score 134.6;   DB 12;   Length 2004;
Best Local Similarity   48.7%;   Pred. No. 4.5e-11;
Matches   489;   Conservative   0;   Mismatches   504;   Indels   12;   Gaps   4;

Qy   265  cagcgcgagcgtggaggagagcggcccggccgtgctgcctggccacggtgaccc      324
Db   280  CAACATGAAGGTGCAGCCCAATCGCAACAAGACCGTCATTTCTCTGTCAATTGGGGACCC   339

Qy   325  gtccgtgttcccggccttccgcacgacgccgtgagcgcgaaacgccgtcgccgcgccgct   384
Db   340  TACTGTGTTGGGAACCTGCCTACAGACCCTACAGAGATGGCCATGAAAGATGCCCT      399

Qy   385  gcgcaccggccagttcaactgctacgctgctgccgtcgcgtcggcccccgccgcacgaagcgc   444
Db   400  GGACTCGGGGAGTACAAATGGCTAGCGCTATGCCCCGTCCATCGGCTACCTATCCAGTCGGGAGGA   459

Qy   445  cgtagcagagagcacttgtcacagggcgtgccctacaagctatcggccgacgagtcttcct   504
Db   460  GGTCGCTTCTTACTACCACCTGTCATGAGGCTCC---TCTGGAAGCTAAGGATGTCATTCT   516

Qy   505  caccgccggcgaactcaggcgatcgaagtcataatcccggtgctggcccagactgccgg   564
Db   517  GACAAGCGGCTGCAGTCAGGCCATTGAGCTATGTCTCTAGCTGTGTTGGCCAATCCT---GG   573

Qy   565  cgccaacatactgcttcccggccaggctatccaaattacgaggcgcgagcggcattcaa   624
Db   574  ACAAAACATCCTCATTCCAAGGCCCGGGTTTTTCCCTCTATAGGACTTGGCTGAGTCTAT   633

Qy   625  caagtggagtccggcacttcgacctcatccccgacaagggtgggagatcgacacatcga   684
Db   634  GGGAATTGAGGTCAAGCTCTACAATCCTCCTGCCTGAGAAGCTTGGGAAATTGACCTAAA   693

Qy   685  ctcggaatccatcgccgacaagagaacaccccgatggtcatcataaaccccaaacaa   744
Db   694  ACAACTGGAATCTCTGATCGATGAAAAAACTGGAATTGCTTGTTGTCAACAACCATCCAA   753

Qy   745  tccgtgccggcagcgttactcctacgaccatctgccaagtgcggcaggaggtggcaaggaa   804
Db   754  TCCCTGCCGTGGCTCCGTTCAGTAAGCGACCACCTTCAGAAGATTTGGCAGTGGCTGAAAG   813

Qy   805  gctcggaatattggtgatcgctgacgaggtttacggcaaactggttctggcgcagcgcccc   864
Db   814  GCAGTGTGTCCCATCTTAGCTGCAGAGATCTATGGTGACATGGTGTTTTCAGATTGCAA   873

Qy   865  gtttatccgatggcgtcttttggcacattgcccggtctttgtccattggatctctgtc   924
Db   874  ATACGAACCACTGGCCAACCTGGCCGCCAGCCACCACCAATGTTCCCATCCGTCCTGTGTGGGGCTGGC   933

Qy   925  caagtcgtggatagtgcgcactggacttgatgggtggtacgaccccccacaaa   984
Db   934  CAAGCGCCTGCCTGGTTGGGGGCTGGGGCTGGGGGCTGGAGGTTGGCGGATCCTCATTCATGGATCGCAAGAGA   993

Qy   985  gatttagagaaactaagatctcttacgctattacctttaatgtctcaacgga   1044
```

```
Query Match        8.3%;   Score 137.4;   DB 8;   Length 103495;
Best Local Similarity   62.2%;   Pred. No. 1.2e-11;
Matches   216;   Conservative   0;   Mismatches   131;   Indels   0;   Gaps   0;

Qy   440  agcgccgtagcagagagcacttgtcacagggcgtgccctacaagctatcggccgacgagctc   499
Db 82305  AGGGCCGGTGGCTGAATATTTAAACGGAGAACTTCCGACGAAGCTGAAGCCCAGGGATGTG   82364

Qy   500  ttcctcaccgccggcggaactcaggcgatcgaagtcatataatcccggtgctgctggcccagact   559
Db 82365  TATATCACCGGAGGATGTAACCAAGCCATAGAGATCGTGATAGATTCTCTTGCCGGAAAT   82424

Qy   560  gccggcgccaacatactgcttcccggccaggctatccaaattacgaggcgcgagcggca   619
Db 82425  CCATCCGCCAACATTCCACTTCCAAGGCCGGGATATCCTCCACTACGATGCCTCGTGCTGTC   82484

Qy   620  ttcaacaagctggagtccggcacttcggacctcatccccgacacaagggtgggagatcgac   679
Db 82485  TATAGCGGCCTTCGAGATTCGCAAATACGATCTTCTCCCGAGAGTGATTGGGAAATCAAT   82544

Qy   680  atcgactcgctggaatccatcgcgcgacaagagacaccaccgccgatgtgtcatcataaaccca   739
Db 82545  CTCGATGGCCTCGAGGCGGCGTGCGGATGAGAATACCGTCGCAATGGTAATCATCAACCCC   82604

Qy   740  aacaatccgtgcggcagcgtttactcctacgaccatctgccaaggt   786
Db 82605  AACAATCCCATGTGGAAACGTCTACACCTACGACCATCTCAACAAGGT   82651
```

RESULT   3
RATTATRA
LOCUS   RATTATRA   2004 bp   mRNA   ROD   15-MAR-1990
DEFINITION Rat tyrosine aminotransferase mRNA, complete cds.
ACCESSION   M18340 K01265
NID   g207155
VERSION   M18340.1 GI:207155
KEYWORDS   tyrosine aminotransferase.
SOURCE   Rat (adult Wistar) hepatocyte, cDNA to mRNA, clones pcTAT[1-4].
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1 (bases 1 to 2004)
AUTHORS   Hargrove,J.L., Scoble,H.A., Mathews,W.R., Baumstark,B.R. and
            Biemann,K.
TITLE   The structure of tyrosine aminotransferase: Evidence for domains
            involved in catalysis and enzyme turnover
JOURNAL   J. Biol. Chem. 264, 45-53 (1989)
MEDLINE   89079691
COMMENT   Draft entry and computer-readable sequence for [1] kindly submitted
            by J.L.Hargrove, 07-NOV-88.
FEATURES   Location/Qualifiers
source   1..2004
            /organism="Rattus norvegicus"
            /db_xref="taxon:10116"

```
Db    994  CATTTTGGCAA--TGAGATTCGAGACGGGCTGGTGAAACTGAGTCAGCGGATCCTGGG 1050
            ||||  ||| ||    |||||| ||| ||||  |||  ||| || ||  ||| |  ||
Qy   1045  cccagcaacctcgttcaggaagcttcctaaaattcttgagaaccacaaaagcagattt 1104
Db   1051  ACCATGCACCATAGTCCAGGGTGCTCTGAAGAGCATCCTTCAGCGAACCCCTCAGGAGTT 1110
            ||||| ||  ||  ||  |||| ||  || ||  ||| || ||| |||||| ||||||||
Qy   1105  ctttaaggaggattattggtctactaaaggaatcatcagagatatgttataggaaataa 1164
Db   1111  CTATCACGACACGTAAGCTTCCTCAAGTCCAATGCGGACCTCTGCTATGGGGCACTGGC 1170
            | || | |  |  || ||  || ||| || ||  |||||| || |  || || | |||
Qy   1165  ggaaaacaaatatattacgtgcctccacaagccagaaggatcgatgtttgtaatggtcaa 1224
Db   1171  TGACCTCCCTGGACTCCAG---CCGGTCCGCCCTTCTCGGAGCCATGTACCTTATGGTGGG 1227
            |  |||| || |  |  ||     || || ||| || ||| || ||||  || | | |||
Qy   1225  actaaacttacatctttggaggagatcctgacacacatagatttt 1269
Db   1228  AATTGAGAGGAGCATTCCCGGAATTCGAGAACGACGTGGGAGTT 1272
```

```
RESULT  4
ATF2009
LOCUS       ATF2009            93695 bp    DNA             PLN       27-OCT-1998
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone F2009 (ESSAII
            project).
ACCESSION   AL021749
NID         g2842474
VERSION     AL021749.1  GI:2842474
KEYWORDS    .
SOURCE      thale cress.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 93695)
AUTHORS     Bevan,M., Koetter,P., Hempel,S., Entian,K.-D., Hoheisel,J.,
            Mewes,H.W., Mayer,K.F.X. and Schueller,C.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 93695)
AUTHORS     EU Arabidopsis sequencing project.
TITLE       Direct Submission
JOURNAL     Submitted (16-OCT-1998) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
FEATURES             Location/Qualifiers
     source          1..93695
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            18..467
                     /gene="F2009.10"
     exon            complement(18..467)
                     /gene="F2009.10"
                     /number=1
     CDS             complement(18..467)
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAA16872.1"
                     /db_xref="PID:e1250050"
                     /db_xref="PID:g2842475"
                     /db_xref="GI:2842475"
                     /db_xref="SPTREMBL:O49442"
                     /translation="MTSYKLFLRVISKRRTWACLFLVVYAILLSSWNSLNSIVNWYG
                     ENHQTSSGLPAIYASVLLGAVFGVLSMAAALFIAVPAIVWISVVVTMAFAGKSRKR
                     VVIEGRKVTKEIAGFVFRVLLKEGNFVALLCALLAYFVFFNSYSSSS"
     exon            complement(2196..2578)
                     /gene="F2009.20"
                     /number=1
     gene            2196..4222
                     /gene="F2009.20"
```

```
     CDS             complement(join(2196..2578,3325..3476,4053..4222))
                     /gene="F2009.20"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAA16873.1"
                     /db_xref="PID:e1250051"
                     /db_xref="PID:g2842476"
                     /db_xref="GI:2842476"
                     /db_xref="SPTREMBL:O49443"
                     /translation="MNSMPVVDRNKSRDAREKRRMETLENMMTERGFFVDVCRCEVLW
                     EEEELCQKRFDCASITNQSLITLVETVRLGDVDSVQHCKYCFTIYHNFTKPDTECY
                     LPFTLVYAVLLSSSRNSLNSIVNWYGENHQTSSGLPAIYASVLLGAVFGVLSMAAALF
                     IAVPAIVVIWISVVVTIAFPVKSRKVVIEGRKVTKEIAGYVFKVLLKEGNFVALLCA
                     VIAYFVFFNSYSYSSSS"
     intron          complement(2579..3324)
                     /gene="F2009.20"
                     /number=1
     exon            complement(3325..3476)
                     /gene="F2009.20"
                     /number=2
     intron          complement(3477..4052)
                     /gene="F2009.20"
                     /number=2
     exon            complement(4053..4222)
                     /gene="F2009.20"
                     /number=3
     exon            4143..4941
                     /gene="F2009.40"
                     /number=1
     gene            4143..6188
                     /gene="F2009.40"
     CDS             join(4143..4941,4987..5044,5096..6188)
                     /gene="F2009.40"
                     /note="similarity to lectin receptor-like serine/threonine
                     kinase lecRK1, Arabidopsis thaliana, PIR2:S68589; Contains
                     Protein kinases signatures and profile; pos. 328-349 &
                     pos. 444-456"
                     /codon_start=1
                     /product="receptor protein kinase like protein"
                     /protein_id="CAA16875.1"
                     /db_xref="PID:e1250053"
                     /db_xref="PID:g2842478"
                     /db_xref="GI:2842478"
                     /db_xref="SPTREMBL:O49445"
                     /translation="MFSKVSILLFSLASLLLFRSTTGIEFIYNSNFTTNTLLLGNAT
                     VKSPPSILTTNQTTFSIGRGLYPSRINASSSASPLPFATSFFISMAPFKHLSPGHG
                     FAFVFLPFSETSAASSQHLGLFNFTNNGDPNSRIFAVEDFVFANOEFNDINDNHVGV
                     DVNSLTSVASETAGFYGGRDGQRFTELKLNSGENYQAWIEFNGSAINVTMARASSRKP
                     IRPLISIPLNLTGVLLDDMFVGFTASTGQLVQSHRILSWSFSNSNFSIGDSVLKSGF
                     IAGVSSGVVLQRLEGDWDWETEYWPHRVQYKDVLEATKGFSDENMIGYGGNSKVYRG
                     VLEGKEVAVKRIMMSPRESVGATSEFLAEVSSLGRLRHKNIVGLKGWSKKGGESLILI
                     YEYMENGSVDKRIFDCNEMLNWEERMRVIRDLASGMYLHEGWETKVLHRDIKSSNVL
                     LDKDMNARVGDFGLAKLQNTSKEMVSTTHVVGTAGYMAPELVKTGRASAQTDVYSFGV
                     FVLEVVCGRRPIEEGREGIVEHIWGLMEKDKVVDGLDERIKANGVFVEEVEMALRIG
                     LLCVHPDPRVRPKMRQVVQILEQGRLVEDGGEREISLLERVKSSYLLETGEGSRQQHP
                     TFQDVWNSSSYSNSFQTYDSILHGR"
     exon            complement(4494..4666)
                     /gene="F2009.40"
                     /number=1
     gene            4494..11254
                     /gene="F2009.30"
     CDS             complement(join(4494..4666,6557..6669,6878..7149,
                     7305..7371,7537..7647,7763..7915,9901..9975,10605..10996,
                     11074..11254))
                     /gene="F2009.30"
                     /note="similarity to ENOD20 gene (small copper-binding
                     protein), Medicago truncatula, PATCHX:E258656"
                     /codon_start=1
                     /product="copper-binding protein-like"
                     /protein_id="CAA16874.1"
                     /db_xref="PID:e1250052"
                     /db_xref="PID:g2842477"
                     /db_xref="GI:2842477"
```

```
          /db_xref="SPTREMBL:O49444"
          /translation="MGLVMRFDLYLMFVMLGLGFTISNGYKFYVGGKDGWVPTPSED
          YSHWSHRNRFQVNDTLHFKYAKGKDSVLEVTEQEYNTCNTHPLTSLSDGDSLFLLSH
          SGSYFFISGNSQNCLKGQKLAVKVLSTVHHSHSPRHTSPSSPVHQELSSPGPSPGVE
          PSSDSNSRVPAPGPATAPNSAGLVGPGMVVLVVNSIIKPIDSFLLKSLPLVVDVAVGI
          STSRQLQASEEPVSSPLSSPALLGSGKEEEQKIIPKRQKVQAVLKSIKQSPKKVNLVA
          ALVRGMRVEDALIQLQVTVKRAAQTYRVIHAARANATHNGLDPDRLLVAEAFVGKG
          LFGKKVAYHAKGRSGIISIPRCRLTVIVRETTPEEEAEIARLKVHNFKKKSKRERQLV
          PHKLIETSPIWNRRGTKANHRSSELVLTIILDVTCVGNMEKNRLDNLTNQNNIYHHNP
          EETCSFRCNGSERIDVDADVVVDVELLVSENIEFNGEDSAIGVTVIGEVEEAEMLR
          "

intron    complement(4667..6566)
          /gene="F2009.30"
          /number=1

intron    4942..4986
          /gene="F2009.40"
          /number=1

          4987..5044
          /gene="F2009.40"
          /number=1

exon      4987..5044
          /gene="F2009.40"
          /number=2

          5045..5095
          /gene="F2009.40"
          /number=2

exon      5096..6188
          /gene="F2009.40"
          /number=3

exon      complement(6567..6669)
          /gene="F2009.30"
          /number=3

intron    complement(6670..6877)
          /gene="F2009.30"
          /number=3

exon      complement(6878..7149)
          /gene="F2009.30"
          /number=3

intron    complement(7150..7304)
          /gene="F2009.30"
          /number=3

exon      complement(7305..7371)
          /gene="F2009.30"
          /number=4

intron    complement(7372..7536)
          /gene="F2009.30"
          /number=4

exon      complement(7537..7647)
          /gene="F2009.30"
          /number=5

intron    complement(7648..7762)
          /gene="F2009.30"
          /number=5

exon      complement(7763..7915)
          /gene="F2009.30"
          /number=6

intron    complement(7916..9900)
          /gene="F2009.30"
          /number=6

exon      complement(9901..9975)
          /gene="F2009.30"
          /number=7

intron    complement(9976..10604)
          /gene="F2009.30"
          /number=7

tRNA      10056..10127
          /gene="F2009.30"
          /note="Lys-tRNA_(TTT)"

exon      complement(10605..10996)
          /gene="F2009.30"
          /number=8

intron    complement(10997..11073)
          /gene="F2009.30"
          /number=8

exon      complement(11074..11254)
          /gene="F2009.30"
```

```
gene      /number=9
          13176..17409

Query Match      8.0%;   Score 133.6;   DB 7;   Length 93695;
Best Local Similarity 63.5%;   Pred. No. 4.1e-11;
Matches 221;   Conservative 0;   Mismatches 124;   Indels 3;   Gaps 1;

Qy   440 agcgccgcagcagcacttgtcacagggcgtgccctacaaagctatcggccgacgacgtc 499
         ||||||| ||||||||| |||| || |||| ||||||  || || |||||| ||||||
Db 33691 AGGGCGGTCGCAGATTACGTGACCATTGACGACCAAGGTAAGCCTAATGATGTA 33750

Qy   500 ttcctcaccgccggcggaactcaggcgatcgaagtcataatccggtgctggcccagact 559
         |||| ||  |||  |||||| || ||||  ||  ||||||  ||| |||| ||| ||
Db 33751 TTCATAACTGTCGATGCAACCAAGGATAGAGTAGTTCTTCAGTCACTGGCTCGACCA 33810

Qy   560 gccggcgccaacatactgcttcccggccaggctatccaaattacgaggcgcagcggca 619
         ||| |||||||| |||||| ||| | |||  ||| ||| |||||  ||||||||||||
Db 33811 AA---CGCTAACACCTTCTCCCAAGGCTCAGTTACCCTCACTACGAGGCTCGTGCCGTC 33867

Qy   620 ttcaacaagctgagtctcgggctcacttcgacctcatcccgacaagggtgggagatcgac 679
         |||  | ||| |||||| || |||| || ||| ||||||  |||||||||| ||||| ||
Db 33868 TACAGTGGACTCGAGGTCCGCAAGTTCGATCTCCTTCCAGAAGAAAGAATGGGAGATTGAT 33927

Qy   680 atcgactcgctggaatccatcgcgcacgaagaacacccggatggtcatcataaaccca 739
         ||| ||| ||| || ||||| ||| |||||| | |||||||| ||||||| |||| |||
Db 33928 CTCCAGCATCGAAGCCATGGCAGATGCAGATGAGAACACAGTTGCAATGGTTATCATAAACCCT 33987

Qy   740 aacaatccgtgcgcagcgtttactcctacgaccatctggccaaggtc 787
         ||||||| || || | ||| ||||||||||| ||||||| |||||||
Db 33988 AACAACCCTGTGGAAATGTTTACTCTTACGATCATCTTAAGAGAGGTC 34035
```

```
RESULT 5
ATAC007048
LOCUS        ATAC007048    61384 bp    DNA          PLN       25-MAR-1999
DEFINITION   Arabidopsis thaliana chromosome II BAC F23N11 genomic sequence,
             complete sequence.
ACCESSION    AC007048
NID          g4512646
VERSION      AC007048.3   GI:4512646
KEYWORDS     HTG.
SOURCE       thale cress.
  ORGANISM   Arabidopsis thaliana
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
             eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
             Arabidopsis.
REFERENCE    1  (bases 1 to 61384)
  AUTHORS    Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., VanAken,S.E.,
             Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
             Carrera,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
             Fraser,C.M. and Venter,J.C.
  TITLE      Arabidopsis thaliana chromosome II BAC F23N11 genomic sequence
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 61384)
  AUTHORS    Lin,X. and Kaul,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (09-MAR-1999) The Institute for Genomic Research, 9712
             Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
REFERENCE    3  (bases 1 to 61384)
  AUTHORS    Lin,X.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-MAR-1999) The Institute for Genomic Research, 9712
             Medical Center Dr., Rockville, MD 20850, USA
COMMENT      On Mar 25, 1999 this sequence version replaced gi:4389529.
             Address all correspondence to:
             Xiaoying Lin
             The Institute for Genomic Research
             9712 Medical Center Dr.
             Rockville, MD 20850, USA
             e-mail: xlin@tigr.org
             BAC clone F23N11 is from Arabidopsis chromosome II and is near the
             molecular marker mi148.
```

The orientation of the sequence is from SP6 to T7 end of the BAC clone.
Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (http://www.tigr.org/tdb/at/at.html).
Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of genomic sequence that are not annotated as genes but have predicted exons by GRAIL are annotated as misc features.

```
FEATURES             Location/Qualifiers
     source          1..61384
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /map="m148"
                     /clone="F23N11"
     misc_feature    1
                     /note="16680 bases before this point were not included in
                     the submitted sequence, but they appear in the overlapping
                     BAC."
     misc_feature    1..13161
                     /note="overlap with BAC clone F5H14
                     (AC006234:116508..129667)."
     repeat_region   complement(69..103)
                     /rpt_family="AT_rich"
     repeat_region   384..409
                     /rpt_family="AT_rich"
     repeat_region   697..721
                     /rpt_family="AT_rich"
     repeat_region   1012..1035
                     /rpt_family="AT_rich"
     misc_feature    1423..1572
                     /rpt_family="AT_rich"
     misc_feature    complement(1528..1620)
                     /note="exon predicted by xgrail, quality excellent"
     mRNA            join(<1719..1934,2049..2277,2589..>3028)
                     /gene="F23N11.1"
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     gene            <1719..>3028
                     /gene="F23N11.1"
                     /note="F5H14.36"
     CDS             join(1719..1934,2049..2277,2589..3028)
                     /gene="F23N11.1"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAD21702.1"
                     /db_xref="PID:g4512647"
                     /db_xref="GI:4512647"
                     /translation="MDRQNSDDIMRFLDGMASSDDVLFGFLDEGNQSPEDFSVNLNAG
                     EDDGGDDDNNNSEDNKAFWQEHEQLQGTLYRTSSIETKIRQATKEALKQVKSKGLY
                     CVCRRPVDGGCRSCLRGEISRHLRDVAGYDCVISKSKWRSCQDIPAGEHEFIEIVDRS
                     GSKKSEMRVVIELSFRAEFEIAKGSEEYKRLISRLPEVYVGKTERLRSLIKILCIAGK
                     KCLRDKKMHMAPWRKHYMQAKWLGTCDRSSSLEASVSEAMEPENWVPVAKPRVSMLN
                     YDGLLGGFSAGPATVAVV"
     repeat_region   complement(2296..2324)
                     /rpt_family="AT_rich"
     repeat_region   complement(2565..2585)
                     /rpt_family="AT_rich"
     misc_feature    complement(3117..3201)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   3327..3389
                     /rpt_family="POLY_A"
     misc_feature    3837..3973
                     /note="exon predicted by xgrail, quality marginal"
     repeat_region   complement(4252..4308)
                     /rpt_family="(TA)n"
     misc_feature    4307..4352
                     /note="exon predicted by xgrail, quality good"
     repeat_region   complement(4870..4893)
                     /rpt_family="AT_rich"
     repeat_region   5283..5408
                     /rpt_family="POLY_A"
     mRNA            complement(<5419..>5724)
                     /gene="F23N11.2"
     gene            complement(<5419..>5724)
                     /gene="F23N11.2"
     CDS             complement(5419..5724)
                     /gene="F23N11.2"
                     /note="predicted by genscan"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD21703.1"
                     /db_xref="PID:g4512648"
                     /db_xref="GI:4512648"
                     /translation="MKLLIFAVIISVVLFPVLVSSRTKCDQLSGKCINGEEKEIMNM
                     RIGLDVSSRRIIQASRIYSEALKKNLPDNRRGEPDQRDNPYRRSCDVHSHCYRFTN"
     misc_feature    5852..5890
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     repeat_region   6449..6487
                     /rpt_family="AT_rich"
     repeat_region   6837..6880
                     /rpt_family="AT_rich"
     repeat_region   7043..7110
                     /rpt_family="POLY_A"
     repeat_region   7231..7292
                     /rpt_family="AT_rich"
     repeat_region   7374..7406
                     /rpt_family="AT_rich"
     repeat_region   complement(7723..7789)
                     /rpt_family="AT_rich"
     misc_feature    complement(7930..8000)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    8174..8302
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    complement(8465..8586)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    8544..8629
                     /note="exon predicted by xgrail, quality good"
     misc_feature    8702..8829
                     /note="exon predicted by xgrail, quality good"
     misc_feature    complement(8934..9057)
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    8936..9051
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    9147..9243
                     /note="exon predicted by xgrail, quality marginal"
     misc_feature    9290..9338
                     /note="exon predicted by xgrail, quality good_shadowexon"
     misc_feature    complement(9317..9461)
                     /note="exon predicted by xgrail, quality marginal"
     misc_feature    complement(9654..9780)
                     /note="exon predicted by xgrail, quality
                     marginal_shadowexon"
     misc_feature    9666..9814
                     /note="exon predicted by xgrail, quality excellent"
     misc_feature    complement(9912..9988)
                     /note="exon predicted by xgrail, quality"
```

```
mRNA            marginal_shadowexon"
                join(<10040. .10154,10340. .10507,10608. .10936,
                11019. .>11237)
                /gene="F23N11.3"
gene            <10040. .>11237
                /gene="F23N11.3"
                /note="F5H14.37; predicted by genscan"
CDS             join(10040. .10154,10340. .10507,10608. .10936,11019. .11237)
                /gene="F23N11.3"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="AAD21704.1"
                /db_xref="PID:g4512649"
                /db_xref="GI:4512649"
                /translation="MTLLVYKGAAKVSIVMIGQQAIMDSYICLLHLTAGIIVESLFNA

Query Match       7.6%;   Score 126.2;  DB 8;  Length 61384;
Best Local Similarity  62.2%;   Pred. No. 5e-10;
Matches 216;   Conservative  0;   Mismatches 128;   Indels 3;   Gaps 1;

Qy   440 agcgcctagcagagacacttgtcacagggcgtgccctacaagctatcggccgacgacgtc 499
         ||||||| ||| ||||||||| |||||||| || ||||| |||||||||||||||  ||||
Db 35491 AGAGCCCTTGCTGATTATATGAACCAGGATCTTCCCGCACAAGTTAACGCCTGAAGATATT 35550

Qy   500 ttcctcaccgccgggcggggaactcaggcgatcgaagtcataatcccggtgctggcccagact 559
         |||||| ||| ||| ||||   ||||||| |||| ||||||| ||||| ||||||||
Db 35551 TTTCTGACGCGGCTGGATCCAACCAAGGATAGAGATCCGTTCGAATCGTTGGCCTCGACCA 35610

Qy   560 gccggcgcccaacatactgcttcccgccaggctcccgctatccaaattacgaggcggcgca 619
         || ||||||| ||  |||||| |||| ||||||||| | |||||| ||||||||| |||||
Db 35611 AA--CGCAAACATCTTGCTCCCACCGTCCTGGCGTTCCTCACTACACGCGTCGTGCTGCT 35667

Qy   620 ttcaacaagctggagtccggcacttgacctcatcccgacaagggtgggagatcgac 679
         || |||||| |||||| |||| |||||||||| |||| |||||||||||||| ||
Db 35668 TACAGTGGTCTCGAGGTTCGCAAGTTTGATCTTCTTCCCGAGAAAGAATGGGAGATTGAT 35727

Qy   680 atcgactcgctggaatccatcgccgacaagaacacacccgcatgtcatccataaaccca 739
         ||| |||| || |||||  ||||||||||||||||  || ||| |||||||| || ||
Db 35728 CTTGAAGGTATCGAAGCACTGAAGCTCACTCCTCACGACGAACACTGTGGCTATGGTTGTAATTAACCCC 35787

Qy   740 aacaatccgtcgcggcagcgtttactcctacgaccatctggccaaggt 786
         ||||| || || |||| ||| ||| |||| |||| ||| ||||||||
Db 35788 AACAATCCCTGTGAAATGTCTACTCTCACGACCATCTCAAAAAGGT 35834


RESULT    6
HSTATR
LOCUS         HSTATR        2754 bp   mRNA             PRI         12-SEP-1993
DEFINITION    Human mRNA for tyrosine aminotransferase (TAT) (EC 2.6.1.5).
ACCESSION     X52520
NID           g36712
VERSION       X52520.1  GI:36712
KEYWORDS      aminotransferase; transferase; tyrosine aminotransferase.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
              Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 2754)
AUTHORS       Scherer,G.
TITLE         Direct Submission
JOURNAL       Submitted (08-MAR-1990) Scherer G., Institute of Human Genetics,
              Albertstr 11, D 7800 Freiburg
REFERENCE     2  (bases 1 to 2754)
AUTHORS       Rettenmeier,R., Natt,E., Zentgraf,H. and Scherer,G.
TITLE         Isolation and characterization of the human tyrosine
              aminotransferase gene
JOURNAL       Nucleic Acids Res. 18 (13), 3853-3861 (1990)
MEDLINE       90326506
COMMENT       See <X52507>-<X52520>.
              Data kindly reviewed (26-JUL-1990) by G. Scherer.
FEATURES             Location/Qualifiers
     source          1. .2754
```

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16q22.1"
                     /tissue_type="placenta (genomic) and liver (cDNA)"
                     /clone_lib="genomic (phage) + cDNA"
                     /clone="lambda-hTAT1, PhcTAT2-16 + phcTAT3a-6"
     CDS             97. .1461
                     /note="tyrosine aminotransferase (AA 1-454)"
                     /codon_start=1
                     /protein_id="CAA36750.1"
                     /db_xref="PID:g36713"
                     /db_xref="GI:36713"
                     /db_xref="SWISS-PROT:P17735"
                     /translation="MDPYMIMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKARWSVR
                     PSDMAKKTFNPIRAIVDNMKVKPNPNKTMISLSIGDPTVFGNLPTDPEVTQAMKDALD
                     SGKYNGYAPSIGFLSREEIASYYHCPEAPLEAKDVILTSGCSQAIDLCLAVLIVNNPS
                     NILVPRPGFSLYKTLAESMGIEVKLYNLLPEKSWEIDLKQLEYLIDEKTACLIVNNPS
                     NPCGSVFSKRHIQKILAVAARQCVPILADEIYGDMVFSDCKYEPLATLSTDVPILSCG
                     GLAKRWLVPGWRLGWLIIHDRRDIFGNEIRDGLVKLSQRIIGPCTIVQGALKSILCRT
                     PGEFYHNTLSFLKSNADLCYGALAAIPGLRPVRPSGAMYLMVGIEMEHFPFENDVEF
                     TERLVAEQSVHCLPATCFEYPNFIRVVITVPEVMLEACSRIQEFCEQHYCAEGSQE
                     ECDK"
     misc_feature    2070. .2075
                     /note="minor mRNA polyadenylation signal"
     polyA_site      2090
                     /note="minor mRNA polyadenylation site"
     repeat_region   2116. .2130
                     /note="direct repeat flanking Alu element"
     misc_feature    complement(2131. .2479)
                     /note="Alu element"
     repeat_region   2480. .2494
                     /note="direct repeat flanking Alu element"
     misc_feature    2738. .2743
                     /note="major mRNA polyadenylation signal"
     polyA_site      2754
                     /note="major mRNA polyadenylation site"
BASE COUNT      689 a     646 c     635 g     784 t
ORIGIN

Query Match       7.5%;   Score 125;  DB 10;  Length 2754;
Best Local Similarity  49.0%;   Pred. No. 1.1e-09;
Matches 422;   Conservative  0;   Mismatches 430;   Indels 9;   Gaps 3;

Qy   297 ccgtgctgccgtgccacgtgacccgtccgtgttccgcacggccgtcg 356
         ||| || ||| |||||||||||| |||| ||||| ||||||||||||
Db   311 CCATGATTCCCTGTCCATTGGGACCCTACTGTGTTTGGAACCTGCCTACAGACCCTG 370

Qy   357 aggccgaagacgacgtcgccgccgggccgcagttcaactgctacggctacgccgccg 416
         ||| |||| || ||||||| ||||||| ||| |||| ||||||| ||||||| ||
Db   371 AAGTTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAATATAATGGCTATGCCCCAT 430

Qy   417 gcgtcggcctcccgccgcaagccgccgtagcagagcactgtcacagggcggtgccct 476
         ||||| ||||| |||||||  |||| || ||| ||||| | || ||||| |||||
Db   431 CCATCGGCTTCCTATCCAGTCGGGAGGAGATTGCTTCTTATTACCACTGTCCTGAGCGAC 490

Qy   477 acaagctatcgccgacgacgtcttcctcaccgcgcggcgaactcaggcgatcgagtca 536
         || | ||| ||||||||||||||  ||||||||||||  |||||||||||||||||
Db   491 CC--CTAGAAGCTAGGACGTCATTCTGACAAGTGGCTGCAGCCAAGCTATTGACCTTT 547

Qy   537 taatcccgtgctggccagactgcgcggcgccaacatactgcttccccggccaggctatc 596
         ||||  || || ||||| ||||||||||||  ||| |||||||| ||||| ||||||
Db   548 GTTTAGCTGTGTTGGGCCAACCCA---GGGCAGAACAATCCTGGTTCCAAGACCTGGTTTCT 604

Qy   597 caaattacgaggcggcggcgccattcaacaagctggaggtccggcacttcgacctcatcc 656
         |||| ||||| ||||| ||| |||||||||| |||||||| |||| ||||||||||||
Db   605 CTCTCTACAAGACTCTGGCTGAGTCTATGGAGTTCAAACTCTACAATTGTTGC 664

Qy   657 ccgacaagggtgggagatcgacatcgactcgctggaatcatccgccgacaagaacacca 716
         |||| ||||| ||||||| ||| ||||||||| | |||  ||||| ||||||||||||
Db   665 CAGAGAAATCTTGGAAATTGACCTGAAAGAATCATCAATTAATGAAAGAAGACAG 724

Qy   717 ccgcgatggtcatcataaacccaaacaatccgtcgcagcgttactcctacgaccatc 776
         || | |||||||||||||||||||||||||||| |||||||| ||||||||||||||
```

```
Db   725  |||     | ||| ||||| || ||| ||  |||||  ||| ||||||
          CTTGTCTCATTGTCAATAATCCATCAAACCCTGTGGGTCAGTGTTCAGCAAACGTCATC  784

Qy   777  tggccaaggtcgcggaggtggcaaggaagctcggaatattggtgatcgctgacgaggttt
          |||     ||| ||  |     ||| | |||  |||||||| || ||| ||| |||||||
Db   785  TTCAGAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCATCTTAGCTGCTGATGAGATCT  844

Qy   837  acggcaaactggttctgggcagcgccccgttatcccgatgggcgtcttgggcacaccgatg  896
          |||   |||  |  |    |||    ||||||||| ||||||||| ||| ||||||||||| |
Db   845  ATGGAGACATGGTGTTTCGGATTGCAAATATGAACCACTGGCCACCCTCAGCACCGATG  904

Qy   897  cccggtcttgtccattgatctctgtccaagtcgtggatagtgcctggatggcgacttg  956
          |||  ||||  ||  ||||| ||||| || ||| || ||||  ||| ||| || ||||
Db   905  TCCCCATCCGTCCTGTCTGGGAGGCGCTGGCCAAGCGGCTGGTTCCTGGCTGGAGGTTGG  964

Qy   957  gatgggtggcggtgtacgacccccacaaagatttagagaaaactaagatctctacgtcta  1016
          |||  ||| ||||  ||   || | |  |  || ||||||| ||  ||||| ||||||||
Db   965  GCTGGATCCTCATTCATGACCGAAGAGACATTTTGGCAA---TGAGATCCGAGATGGGC  1021

Qy  1017  ttacgaattaccttaatgtctcaacggaccagcaaccttcgttcaggaagctcttccta  1076
          |  |||||  |||  ||||| |  ||  || |  ||||| |||  ||||||| || ||||
Db  1022  TGGTGAAGCTGAGTGAGCCATTTGGGACCCTGTACCATTGTCCAGGGAGCTCTGAAAA  1081

Qy  1077  aaattcttgagaacaacaaaagcagagattcttttaagaggagattattggttactaaaggaat  1136
          |  |||||||||| || ||  |||| | || |||| || ||||| ||||| || ||||||||
Db  1082  GCATCCTATGTCGCACCCCGGGAGAGTTTTACCAACACTCTGAGCTTCCTCAAGTCCA  1141

Qy  1137  catcagagatatgttataggg  1157
          |||||| |||| |||  ||||
Db  1142  ATGCTGATCTCTGTTATGGGG  1162
```

```
RESULT  7
HSTYAT

LOCUS        HSTYAT      2051 bp    mRNA        PRI        13-JUL-1995
DEFINITION   H.sapiens mRNA for tyrosine aminotransferase.
ACCESSION    X55675
NID          g37501
VERSION      X55675.1  GI:37501
KEYWORDS     tyrosine aminotransferase.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
             Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2051)
  AUTHORS    Labrie.F.
  TITLE      Direct Submission
  JOURNAL    Submitted (26-SEP-1990) Labrie F., CHUL Research Centre, Laboratory
             of Molecular Endocrinology, 2705 Blvd Laurier, Sainte-Foy Quebec
             G1V 4G2, Canada
REFERENCE    2  (bases 1 to 2051)
  AUTHORS    Seralini,G.E., Luu-The,V. and Labrie,F.
  TITLE      Cloning and expression of human tyrosine aminotransferase cDNA
  JOURNAL    Biochim. Biophys. Acta 1260 (1), 97-101 (1995)
  MEDLINE    95092801
FEATURES             Location/Qualifiers
     source          1..2051
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /dev_stage="adult"
                     /tissue_type="liver"
                     /clone_lib="# HL1001b. Lot # 2102"
                     /clone="B1"
                     /map="16q 22.1"
     CDS             51..1415
                     /EC_number="2.6.1.5"
                     /codon_start=1
                     /product="tyrosine aminotransferase"
                     /protein_id="CAA39210.1"
                     /db_xref="PID:g37502"
                     /db_xref="GI:37502"
                     /db_xref="SWISS-PROT:P17735"
```

```
/translation="MDPYMIQMSSKGNLPSILDVHVNVGGRSSVPGKMKGRKARWSVR
PSDMAKKTFNPIRAIVDNMKVKPNPNKTMISLSIGDPTVFGNLPTDPEVTQAMKDALD
SGKYNGYAPSIGFLSSREEIASYYHCPEAPLEAKDVILTSGCSQAIDLCLAVLANPGQ
NILVPRPGFSLYKTLAESMGIEVKLYNLLPEKSWEIDLQLEYLIDEKTACLIVNNPS
NPCGSVFSRKRHLQKILAVARQCVPILADEIYGDMVFSDCKYEPLATLSTDVPILSCG
GLAKRWLVPGWRLGWILIHDRRDIFGNEIRDGLVKLSQRILGPCTIVQGALKSILCRT
PGEFYHNTLSFLKSNADLCYGALAAIPGLRPVRPSGAMLMVGIEMEHFPEFENDVEF
TERLVAEQSVHCLPATCFEXPNFIRVVITVPEVMLEACSRIQEFCEQHYHCAEGSQE
ECDK"
```

```
BASE         polyA_signal   531 a    493 c    496 g    531 t
  ORIGIN                     2021..2026

Query Match                7.5%;    Score 125;  DB 10;   Length 2051;
Best Local Similarity     49.0%;    Pred. No. 1.1e-09;
  Matches  422;  Conservative    0;  Mismatches   430;   Indels    9;   Gaps    3:

Qy   297  ccgtgctgcgctggcccacgtgacccgtccggtgttccggcccgttccgcacggccgtcg  356
          | ||||||||||||    ||  |||| | ||||||  || ||||   |||| ||| | | |
Db   265  CCATCGATTTCCCTGTCCATTGGGACCCTACTGTGTTGGAAACCGCCTACAGACCCTG  324

Qy   357  aggccgaagacgccgtcgcgccgcgctgcgcaccggcagttcaactgctacggccgccg  416
          ||||  |||||||| | | ||||| ||| |||||||| || |||| | || |||| ||||
Db   325  AAGTTACCCAGGCAATGAAAGATGCCCTGGACTCGGGCAAAATAATATGGCTATGCCCCAT  384

Qy   417  gcgtcggcctcccccgccgcgcacgaagcgcctagcacaggcgggcgtgccct  476
          ||  ||||   ||| |   || || |||| | ||||  || ||| | ||| ||
Db   385  CCATCGGCTTCCTATCCAGTGCGGGAGGAGATTGCTTCTTATTACCACTGTCCTGAGGCAC  444

Qy   477  acaagctatcggccgacgacgtcttcctcaccgcgcggaactcaggcgatcgaagtca  536
          || | |       |  |||  || ||||   ||||  || |||   |||||  |||  |||
Db   445  CC---CTAGAAGCTAAGGACGTCATTCTGACAAGTGGCTCGCAGCCAAGCTATTGACCTTT  501

Qy   537  taatcccggtgctgctgccagactgcgcgggccaacatactgcttccccggcccaggctatc  596
          |  || ||||| ||||| || |  |   |  ||  ||  |   ||  ||| || |||| ||
Db   502  GTTTAGCTGTGTTGGCCACCCA---GGCCAAAACATCCTGGTTCCAAGACCCTGGTTTCT  558

Qy   597  caaattacgagcgcgagcggcattcaacagctggaggtccggcacttcgacctcatcc  656
          | |||  || ||| |||| |   |||| | |||  |||| || |   ||| || |||||
Db   559  CTCTCTACAAGACTCTGGCTGAGTCTATGGGAGTCCTGAGTGAGGTCAAACTCTACAATTGTTGC  618

Qy   657  ccgacaaggggtggagatcgacatcgactcgctggaatccatcgcgccgcaagaacacca  716
          || ||  ||||| |||||| ||  |  || |||| | |||  |||  || ||| || | |
Db   619  CAGAGAAATCTTGGGAAATTGACCTGAAACAACTGAATATCTAATTGATGAAAAAGACAG  678

Qy   717  cccgcgatgtgtcatcataaaccaacaatccgtgcggcagcgttactcctacgaccatc  776
          ||| | ||  || |   |||| | |  | |||| | |||| || ||| | || || ||
Db   679  CTTGTCATTGTCAATATCCATCAATATGTCAATAAACCCCTGTGGGTCAGTGTTCAGCAAACGTCATC  738

Qy   777  tggccaaggtcgcggaggtggcaaggaagctcggaatattggtgatcgctgacgaggttt  836
          ||| | |||| |  ||  |    || | |||| ||||||||| ||  |||| |||||||||
Db   739  TTCAGAAGATTCTGGCAGTGGCTGCACGGCAGTGTGTCCCCATCTTAGCTGCTGATGAGATCT  798

Qy   837  acggcaaactggttctgggcagcgccccgtttatcccgatgggcgtcttgggcacattg  896
          |||   | ||  ||    ||| |   |  ||| ||||||||| ||| || ||||| | ||
Db   799  ATGGAGACATGGTGTTTCGGATTGCAAATATGAACCACTGGCCACCCTCAGCACCGATG  858

Qy   897  cccggtcgtctcattgatctctgtccaagtcgtggatagtgcctggatggcgacttg  956
          |||  ||| ||||| | ||||| ||||||| | || | || || || || ||| ||||
Db   859  TCCCCATCCGTCCTGTCCTGTGGGAGGCTGGCCAAGCGGCTGGTTCCTGGCTGGAGGTTGG  918

Qy   957  gatgggtggcggtgtacgacccccacaaagatttagagaaaactaagatctctacgtcta  1016
          |||  ||| ||||  ||   || | |  |  || ||||||| | |  |||| |||||||||
Db   919  GCTGGATCCTCATTCATGACCGAAGAGACATTTTGGCAA---TGAGATCCGAGATGGGC  975

Qy  1017  ttacgaattaccttaatgtctcaacggaccagcaaccttcgttcaggaagctcttccta  1076
          |  |||  |||  |||||| |  || || |  |||||| ||| |||||||||| || ||||
Db   976  TGGTGAAGCTGAGTGAGCCATTTGGGACCCTGTACCATTGTCCAGGGAGCTCTGAAAA  1035

Qy  1077  aaattcttgagaacaacaaaagcagagattcttttaagaggagattattggttactaaaggaat  1136
          |  |||||||||| || || ||| || || ||||| || ||||| ||||| || ||||||||
Db  1036  GCATCCTATGTCGCACCCCGGGAGAGTTTTACCAACACTCTGAGCTTCCTCAAGTCCA  1095
```

Qy   1137 catcagagatatgttataggg 1157
          ||| ||||||| ||||||| |||
Db   1096 ATGCTGATCTCTGTTATGGGG 1116

RESULT   8
LOCUS   AB007644
ACCESSION   AB007644   73840 bp   DNA   PLN   13-FEB-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone: K19P17,
            complete sequence.
ACCESSION   AB007644
NID   g2564044
VERSION   AB007644.1 GI:2564044
KEYWORDS   HTG.
SOURCE   Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
            clone:K19P17.
ORGANISM   Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1 (sites)
AUTHORS   Nakamura,Y., Sato,S., Kaneko,T., Kotani,H., Asamizu,E., Miyajima,N.
            and Tabata.S.
TITLE   Structural analysis of Arabidopsis thaliana chromosome 5. III.
            Sequence features of the regions of 1,191,918 bp covered by
            seventeen physically assigned P1 clones
JOURNAL   DNA Res. 4 (6), 401-414 (1997)
MEDLINE   98162728
REFERENCE   2 (bases 1 to 73840)
AUTHORS   Nakamura,Y.
TITLE   Direct Submission
JOURNAL   Submitted (03-OCT-1997) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES   Location/Qualifiers
source   1..73840
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /chromosome="5"
                /clone="K19P17"
                /clone_lib="Mitsui TAC"
BASE COUNT   23478 a 13438 c 12939 g 23985 t
ORIGIN

Query Match   7.1%; Score 118.2; DB 7; Length 73840;
Best Local Similarity 60.8%; Pred. No. 7e-09;
Matches   211; Conservative   0; Mismatches 133; Indels   3; Gaps   1;

Qy   440 agcgccgtagcagagcacttgtcacaggcgtgccctacaagctatcggccgacgacgtc 499
          |||||| | |||||||| || || |||| || |||||||| || ||||| ||||||||
Db 50877 AGGGCAATAGCAGAGTATCTATCGCGTGATCTTCCATACAAACTTCACAGGATGATGTG 50936

Qy   500 ttcctcaccgcggaactcaggcgagtgaagtcataatcccgtgctgcgcccagact 559
          || |||| || | |||| | ||||||| | |||||| | ||||| | |||||||
Db 50937 TTTATCACATCGGGTTGCACGCAAGCGATCGATGTAGCATTGTCCATGTTAGCTCG---T 50993

Qy   560 gccggcgccaacatactgttcccggcaggctatccaaattacgagcgcgagcgagggca 619
          |||||||||| ||| || || |||| ||| | |||| |||| ||| ||||
Db 50994 CCCAGGCGCTAATATACTTCTTCCAAGGCCTGGTTTCCAACTCATGAACTCGTGCTAAG 51053

Qy   620 ttcaacaagctggagtccggcacttcgacctcatccccgacaagggtggagatcgac 679
          || ||||| ||||||||||||| || |||| ||||| |||| |||||||||||||
Db 51054 TTTAGACACCTTGAAGTTCGCTACGTCGCATCTTCTTCCGGAAAATGGATGGGAGATCGAT 51113

Qy   680 atcgactcgctggaatccatcgccgacaagaacaccacccgatggtcatcataaaccca 739
          || |||| ||| ||| ||| ||||| || || ||||||||| ||| ||| |||||||
Db 51114 CTTGATGCTGTCGCAGGCTCTTGCAGACGAAAAACACGGTTGCTTTGGTTGTTATAAACCT 51173

Qy   740 aacaatccgtgcggcgcagcgttactcctacgacacattcggccaaggt 786
          ||||| ||||| ||| ||||||||| ||||| |||| ||| ||||||||
Db 51174 GGTAATCCTGCGGGATGTCTATAAGCTACCAGCACATTTGATGAAGGT 51220

RESULT   9
LOCUS   SCH10
DEFINITION Streptomyces coelicolor cosmid H10.
ACCESSION   SCH10   39524 bp   DNA   BCT   04-MAY-1999
NID   AL049754
VERSION   g4753846
KEYWORDS   AL049754.1 GI:4753846
SOURCE   ABC transporter; adenylosuccinate synthetase; aspartate
            aminotransferase; ATP-dependent protease ATP-binding subunit;
            cytochrome P-450 hydroxylase; fba; fructose 1,6-bisphosphate
            aldolase; lipase/esterase; oxidoreductase; purA; RNA polymerase
            sigma factor; spermidine synthase; transmembrane efflux protein;
            two-component regulator; two-component sensor kinase; uridine
            5'-monophosphate synthase.
ORGANISM   Streptomyces coelicolor.
            Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1 (bases 1 to 39524)
AUTHORS   Redenbach.M., Kieser.H.M., Denapaite.D., Eichner.A., Cullum.J.,
            Kinashi.H. and Hopwood.D.A.
TITLE   A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL   Mol.Microbiol. 21 (1), 77-96 (1996)
MEDLINE   97000351
REFERENCE   2 (bases 1 to 39524)
AUTHORS   Murphy.L. and Harris.D.
JOURNAL   Unpublished
REFERENCE   3 (bases 1 to 39524)
AUTHORS   James.K.D., Parkhill.J., Barrell.B.G. and Rajandream.M.A.
TITLE   Direct Submission
JOURNAL   Submitted (05-MAY-1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT   Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7_01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.JP/
            jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
            correct initiation codon. Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon). If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone. It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.

```
          Cosmid H10 lies between 66T3 and H44 on the AseI-H genomic
          restriction fragment.
FEATURES           Location/Qualifiers
     source        1..39524
                   /organism="Streptomyces coelicolor"
                   /strain="A3(2)"
                   /db_xref="taxon:1902"
                   /clone="Cosmid H10"
     gene          1..360
                   /gene="SCH10.01"
     misc_feature  1..44
                   /gene="SCH10.01"
                   /note="overlap with S.coelicolor cosmid St6T3 from 35058
                   35101"
     CDS           <1..360
                   /gene="SCH10.01"
                   /note="SCH10.01, hypothetical protein, partial CDS, len:
                   >119 aa; probable CDS suggested by positional base
                   preference, GC frame analysis and amino acid composition"
                   /codon_start=1
                   /transl_table=11
                   /label=SCH10.01
                   /product="hypothetical protein"
                   /protein_id="CAB42010.1"
                   /db_xref="PID:e1428991"
                   /db_xref="PID:g4753847"
                   /db_xref="GI:4753847"
                   /translation="DPRIAAALAGVRRGEEAQRAAAAGDLTVLIGYDLRYWRELATL
                   FGNPYLCDFLHRLRVQSWVCTVQHLRRLSELRGALWSGHTALVDALARRDVPGARALV
                   DAYNSHSLALIEGLAGE"
     gene          462..1127
                   /gene="SCH10.02"
     CDS           462..1127
                   /gene="SCH10.02"
                   /note="SCH10.02, hypothetical protein, len: 221 aa;
                   unknown function, probable CDS suggested by positional
                   base preference, GC frame analysis and amino acid
                   composition"
                   /codon_start=1
                   /transl_table=11
                   /label=SCH10.02
                   /product="hypothetical protein"
                   /protein_id="CAB42011.1"
                   /db_xref="PID:e1428992"
                   /db_xref="PID:g4753848"
                   /db_xref="GI:4753848"
                   /translation="MRSPLLACDLWLVPLVDVLCHTPDNPFAEELAQYDKVLAEAGLP
                   PVPVYQYMPGLSGEVAPVAGEFDYDALHFLRRAHLLQVCGLPVTPVDELGGDYEQLLEM
                   FESTAQQSHLVWHYDHAGAYVPVPDFPHPLSSDELLAGGGPLGSSHTLLRELEAVAPAL
                   GIDPANPPAPPQPPLAPTELEEPAVPAPHDSSPFARERHVWLGLHAAATRSLAQGSMI
                   VFS"
     gene          complement(1143..1433)
                   /gene="SCH10.03c"
     CDS           complement(1143..1433)
                   /gene="SCH10.03c"
                   /note="SCH10.03c, hypothetical protein, len: 96 aa;
                   unknown function, probable CDS suggested by positional
                   base preference, GC frame analysis and amino acid
                   composition"
                   /codon_start=1
                   /transl_table=11
                   /label=SCH10.03c
                   /product="hypothetical protein"
                   /protein_id="CAB42012.1"
                   /db_xref="PID:e1428993"
                   /db_xref="PID:g4753849"
                   /db_xref="GI:4753849"
                   /translation="MRVLLKATLDTEKSNEVIRSGRMEALIKETVDRIGPEAAYFGTI
                   DGRRTALLVFDMQDSSQMPPIGEPFLMELNAEVDLCPIMNGDDLAKGLSQLG"
     gene          complement(1472..2356)
                   /gene="SCH10.04c"
     CDS           complement(1472..2356)
                   /gene="SCH10.04c"
```

```
                   /note="SCH10.04c, probable RNA polymer.. sigma factor,
                   len: 294 aa; similar to many e.g. TR:O05735 (EMBL:U87308),
                   sigE, Mycobacterium avium extracytoplasmic function
                   alternative sigma factor (251 aa), fasta scores: opt: 226
                   z-score: 266.2 E(): 1.6e-07, 29.8% identity in 168 aa
                   overlap"
                   /codon_start=1
                   /transl_table=11
                   /label=SCH10.04c
                   /product="putative RNA polymerase sigma factor"
                   /protein_id="CAB42013.1"
                   /db_xref="PID:e1428994"
                   /db_xref="PID:g4753850"
                   /db_xref="GI:4753850"
                   /translation="MPARSDDMARQFELDRPRLRAVAYRILGSLGEADDALQEAWLRA
                   DRANTSEVGNPSGWLTTVARVCNLLRARDTREEPLDDAARRQPAATGTAADPAEE
                   VQLADEVGIALLIVLDTLGPAERLAFVLHDMFDVPFDDIAAMLDKTPAATRQLASRAR
                   RRVRCVPAPEADLPRRRAVDAYLAATRGGDFDALVALLHPDVVLSADAAVVPTEPPV
                   TVSGMEQVARGAMASMARSAAVVLVDGMVGMAMAENGRLRVVLRFDFAADGRITGI
                   DVIAEPARLNELDITGIG"
     gene          complement(2472..2672)
                   /gene="SCH10.05c"
     CDS           complement(2472..2672)
                   /gene="SCH10.05c"
                   /note="SCH10.05c, hypothetical protein, len: 66 aa;
                   unknown function, probable CDS suggested by positional
                   base preference and GC frame analysis"
                   /codon_start=1
                   /transl_table=11
                   /label=SCH10.05c
                   /product="hypothetical protein"
                   /protein_id="CAB42014.1"
                   /db_xref="PID:e1428995"
                   /db_xref="PID:g4753851"
                   /db_xref="GI:4753851"
                   /translation="MTDRYSAASCCGPYGGENGPEDCGDPVRFEVARHLRAPLRVCPV
                   HLGPSLLATGVLWPPGIVLVR"
     gene          complement(2749..3240)
                   /gene="SCH10.06c"
     CDS           complement(2749..3240)
                   /gene="SCH10.06c"
                   /note="SCH10.06c, hypothetical protein, len: 163 aa;
                   unknown function, probable CDS suggested by positional
                   base preference, GC frame analysis and amino acid
                   composition"
                   /codon_start=1
                   /transl_table=11
                   /label=SCH10.06c
                   /product="hypothetical protein"
                   /protein_id="CAB42015.1"
                   /db_xref="PID:e1428996"
                   /db_xref="PID:g4753852"
                   /db_xref="GI:4753852"
                   /translation="MRKPTTRTWIWLAAAVVVTGGAFYVDGHRVSAAPHAGDPKCDE
                   IVSRLPDDIPGASRDWALGDGVASWGGQKAVFRCGAEELQPNVNLCVTADGVDWVLDE
                   ARLKRDGVSVLRTYGRSPAVEFTYSGPREVGGILTALDPAVKKWIPQERKCIGLDDAA
                   TVL"
     gene          complement(3412..4695)
                   /gene="SCH10.07c"
     CDS           complement(3412..4695)
                   /gene="SCH10.07c"
                   /note="SCH10.07c, purA, probable adenylosuccinate
```

```
Query Match             6.6%;   Score 110.2;  DB 1;  Length 39524;
Best Local Similarity  49.8%;   Pred. No. 1.1e-07;
Matches   308;  Conservative   0;  Mismatches  308;   Indels   3;   Gaps   1;

Qy   270  cgagcgtgaggagcgggcgcgcgccgtgccgctgccgcccacggtgacccgtccg 329
          ||||||| |||||||||||| | | ||| | ||||||| |||  |  |||||||
Db 33265  CGATCTTCCGCGAGATGTCCGCCCTGCCGGCGACCACGGGCGATCAACCTCGGCCAGG 33324

Qy   330  tgttcccggccttccgcacggcgtcgaggccgaagacgacccgtcgccgccgctgcgca 389
          ||||||||| |  |    ||| | | || | |||| ||  ||||| |  |||||||
```

```
Db   33325  GCTTCCCGACACCGACGGCCCCGAGGAGGTCCGGGAGGCCGGGTGCGGGCACTGCGCG 33384
              | |      |||     ||| ||| |     | ||     | || |||| |   | |
Qy    390  cggccagttcaactgctacggcgcggcggtcggcctcccccgccgacggaggccgtag 449
              ||   |||    |||  || || || ||||   ||       ||||||  |||||||
Db   33385  ACGGGCCGGCGGCAACCAGTACCCCCCGGCCCGGCGGCGTCCCCGGCGTCGCGCCGCCGGTCG 33444

Qy    450  cagagcacttgtcacaggg---cgtgcctacaagctatcggccgacgacgtcttcctca 506
              |  ||| |||  || ||||        || || ||  ||||   ||| ||| |||  |||
Db   33445  CCGGCCACCAGCAGCGCCGCGCGCTACGGACCGCCTGTCTCACGACCCCGACACCACCGAGGTCCTGGTCA 33504

Qy    507  ccgccgcgggactcaggcgatcgaagtcataatcccggtgctggccagactgccggcg 566
              ||||| ||      ||  || ||  ||||||  ||||  ||| ||   |||  | ||||
Db   33505  CCGCCGGCCCACGGAGGCCATCGGCGGCCCGCTGTCGGCCCTGTCGGAGCCCGGGACG 33564

Qy    567  caaacactgcttcccgccaggctcatccaaattacgagggcgcgagcggcattcaaca 626
              |       |||     |||   ||| || |  ||| ||||    | ||   |||
Db   33565  AGGTCGTCGCGCGGCTGGAGCCGTACTACGACTCGTACGCGGCGCCTGCCATCGCGATGGGCGGCG 33624

Qy    627  agctggaggtccggcacttgacctcgacctcatccccgacaaagggtgggagatcgacatcgact 686
              |||  |  ||     ||||| ||     ||  || |  ||  || |       ||  ||||| |
Db   33625  GCACCCCGCCGTGCCGGTGACCCTGCGGCCCGCCCTGGCCCTGTGGCCGCACGAGGGCACGTGGACCTGGACG 33684

Qy    687  cgctggaatccatcgccgacaacaagaacaccaccgcgatggtcatcataaaccccaaacaatc 746
              ||           || ||     | || |   ||  | || ||  ||  | |  |||||| |
Db   33685  AGCTGCCGCCGCACGCGGTCACCGACCATCGCTCCGCTGCTGCTGCTGACCACCCCGCACACCCC 33744

Qy    747  cgtgcgcagcgttactcctacgaccatctgccaaggtcgcggagtggcaaggaagc 806
              ||  |||  |     ||  |  ||   ||| ||    ||  ||  ||| | |||  |
Db   33745  CGACGGCGACGGTCCTGGGGTGTCCTCACCCGCCGACGGCTGCGAGCTCGGCGGTGGAGC 33804

Qy    807  tcggaatattggtgatcgctgacgaggtttacggcaaactggttctggcgacgcgcccgt 866
              ||       ||  || | ||||  ||  ||  ||||| | |||| || | |
Db   33805  GGGACCTGCTGGTGGTGACCGACGGAGGTCTACGAGCGGCTACCGACCGTCTTCGGCCACCACCTGGAGC 33864

Qy    867  ttatcccgatgggcgtctt 885
              | ||| || ||||| |||
Db   33865  ACATCCCCTGGCGTCCTT 33883
```

```
RESULT 10
AB018112/c
LOCUS       AB018112      70475 bp   DNA                PLN      20-FEB-1999
DEFINITION  Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MAB16,
            complete sequence.
ACCESSION   AB018112
NID         g3702730
VERSION     AB018112.1  GI:3702730
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
            clone:MAB16.
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosida; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (sites)
AUTHORS     Nakamura,Y.
TITLE       Structural Analysis of Arabidopsis thaliana Chromosome 5. IX
JOURNAL     Unpublished (1998)
REFERENCE   2  (bases 1 to 70475)
AUTHORS     Nakamura,Y.
TITLE       Direct Submission
JOURNAL     Submitted (06-OCT-1998) to the DDBJ/EMBL/GenBank databases.
            Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..70475
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="5"
```

```
                                         /clone="MAB16"
                                         /clone_lib="Mitsui P1"
BASE COUNT
ORIGIN      23859 a  11672 c  11571 g  23373 t

Query Match            6.8%;  Score 110.2;  DB 7;  Length 70475;
Best Local Similarity  59.48%;  Pred. No. 1e-07;
Matches 206;  Conservative  0;  Mismatches 138;  Indels  3;  Gaps  1;

Qy    440  agcgccgtagcagagcacttgtcacaggcgtgccctacaagctatcggccgacgacgtc 499
              || | |||  ||| |      ||  ||||   |||| ||| |  |||| ||||| || |
Db   59787  AGGGCGTTGCAGAGTACCTCTCCTAGTGATCTCTTTACCAGATATCACCAAATGATGTT 59728

Qy    500  ttcctcaccgccggaactcaggcgatcgaagtcataatcccggtgctgcccagact 559
              ||||  || |    |    || || | || |||| |  |  ||    | ||  ||||
Db   59727  CATATCACAGCTGGTTGTGTGCAAGCCATTGAGATTTGATATCCGCTCCGC---AATT 59671

Qy    560  gccgggccaacatactgcttccccggccaggctatccaaattacgagcggcgagcggca 619
              |||  || | | |  ||||     |||   || |   |  |||||| |  | |||  | |
Db   59670  CCTGGAGCTAACATTCTGCTACTACCCGGATGTACGAGTTCCCAGCGGCT 59611

Qy    620  ttcaacaagctggaggtccggcacttcgacctcatccccgacaaggggtgggagatcgac 679
              |  |  ||  ||| |||| | || |||   | ||| || ||  ||  || ||| ||| ||
Db   59610  TTTTGCCAACTCGAGGTTCGTTATTCGACCTTCCTACCCGGAAAATGGTTGGGATGTTGAT 59551

Qy    680  atcgactgctgaatccatcgccgacagaacaccgcgatggtcatcataaaccca 739
              || |  | ||| | || | || |  || |   ||  | ||| || | |  |||||
Db   59550  CTAGATGGAGTTGAAGCTGTTGCGGGATGGATGACAAAAAACCGTTGCTATACTTGTTATCAACCCT 59491

Qy    740  aacaatccgtgcgcagcgttactcctacgaccatctggccaaggt 786
              || |  |||  |  | |  |||  |  || ||| ||  ||  |||||
Db   59490  TGCAATCCATGTGGGAATGTTTTCTCTCGCCAACATCTTCAAAAGGT 59444

RESULT 11
SC2H4/c
LOCUS       SC2H4       25970 bp   DNA               BCT      07-SEP-1998
DEFINITION  Streptomyces coelicolor cosmid 2H4.
ACCESSION   AL031514
NID         g3559956
VERSION     AL031514.1  GI:3559956
KEYWORDS    aminotransferase; beta-mannosidase; permease; secreted
            beta-mannosidase; sugar transport system; sugar-binding
            lipoprotein.
SOURCE      Streptomyces coelicolor.
ORGANISM    Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteridae; Actinobacteria;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 25970)
AUTHORS     Oliver,K. and Harris,D.
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 25970)
AUTHORS     Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE       Direct Submission
JOURNAL     Submitted (01-SEP-1998) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
REFERENCE   3  (bases 1 to 25970)
AUTHORS     Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
TITLE       A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL     Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE     97000351
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC.
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
```

```
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon. If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 2H4 lies
between 2G5 and AH10 on the AseI-B genomic restriction fragment.

FEATURES             Location/Qualifiers
     source          1..25970
                     /organism="Streptomyces coelicolor"
                     /strain="A3(2)"
                     /db_xref="taxon:1902"
                     /clone="cosmid 2H4"
     gene            1..1822
                     /gene="SC2H4.01"
     CDS             <1..1822
                     /gene="SC2H4.01"
                     /note="SC2H4.01, probable ATP/GTP binding protein, partial
                     CDS, len: >606 aa; contains PS00017 ATP /GTP-binding site
                     motif A (P-loop)"
                     /codon_start=2
                     /transl_table=11
                     /product="putative ATP/GTP binding protein"
                     /protein_id="CAA20595.1"
                     /db_xref="PID:e1319219"
                     /db_xref="PID:g3559957"
                     /db_xref="GI:3559957"
                     /db_xref="SPTREMBL:086584"
                     /translation="IDSTAPKAPVITSLSPVTQCTADLCEGKGGPTAGSFTFQPNAA
                     DTDITGYRWRLLTTSAKEAKAVTGKSVTKSVTTVSDVTPSLSGTQVLSVEAKDVRSRWGAPAE
                     FSFKVAPAGATGTWHFDDALPSGVTIAKDTAAEGTRHDATLYTAGAGWSTLARRGD
                     ADNSLWLNDGSDTTQRAGYAATSAPAVNTRDSFTVSAWAYLTDTSQTRVVMAAPGTYG
                     SAFTLYYSASYKKWVFNRTAGDVKDKPVYLRSLGDATAPPLKVWTHLAAVFDTKKDTN
                     KANDTIQLFVNGRPGQGEPVVLDGVSTAYQPWVSSGGLQIGRSLVGGAWGEHFRGRLDE
                     VNVWQRVLTPDEITQAQLLEEGVPANELVAQWDAASSTGTTVKELSSYPAPSLTLSA
                     AGAVLDEENNALVLDGSAGYASAIGPVVDETGSFTVSARVQLDSAKLASKPVGYEALV
                     AGQQASAGESSWALWAVKPADGVYQWKFTRTAVGADGKVTQSAEAPGGDIAETDTWVQ
                     VTGVFDAQEPWEWTDPADSAKTETRYGRLHLYVGEFDQPSETASGFTTPQYGAGALAV
                     GRGSRGGTTGHHLPGALEELRVWTGAMSADQVRSQVLGGV"
     misc_feature    188..211
                     /gene="SC2H4.01"
                     /note="PS00017 ATP/GTP-binding site motif A (P-loop)"
     RBS             1949..1953
                     /note="possible RBS upstream of SC2H4.02"
     gene            1959..8510
                     /gene="SC2H4.02"
     CDS             1959..8510
                     /gene="SC2H4.02"
                     /note="SC2H4.02, unknown, len: 2183; very limited
                     similarity to proteins of the RHS family e.g. WAPA_BACSU
                     wall-associated protein precursor (2334 aa), fasta scores;
                     opt: 141 z-score: 296.8 E(): 2.7e-09, 23.4% identity in
                     1237 aa overlap, and RHSA_ECOLI RHSA protein precursor
                     (1377 aa), fasta scores; opt: 135 z-score: 222.4 E():
                     3.8e-05, 23.0% identity in 1016 aa overlap. Contains
                     possible membrane anchor around aa 1900. Alternative start
                     at aa 27 would give N-terminal signal sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SC2H4.02"
                     /protein_id="CAA20596.1"
                     /db_xref="PID:e1319220"
                     /db_xref="PID:g3559958"
                     /db_xref="GI:3559958"
                     /db_xref="SPTREMBL:086585"
                     /translation="MRHHPLGVPGAGTGPGPHRRRAVRPMRRVSAVVGIALLPGLLA
                     PAATADDVDPLGRPELSTPRAQHVAPFKARPDARTAAVVRDSARADQSAAGRAVKDRS
                     RGTSWPTAGSARLTASGGAADEAAPGRLPVTLAPPAAKKAKRADRVKVDVLDQKAARR
                     LGVKGVVLKVTGPADGGAARLGIDYSAFAAAYGGDWAGRLQVLRLPDCTTACPPATPV
                     KFTNERGDERLTADLAFAPKAASQQTMTLALAAGTASADGDYKATPLAPSSTWEAGGS
                     SGSFTWSYPLKAPEPAAGEAPKLEISYDGSGVDGRTSTSNNQGTEVGEGFDITSSFIE
                     RKYGSCTDDGHADRRDLCWYDNASLVLNGKATELVKDDTTGQWRLKNDDASKVTHST
                     GADNGDDNGEYWTVTGDGTRFVPGQNKLDGAAADVRTKSVWTVPVPFGDDEGEPGYAD
                     GTGFTSRDEKQAWRWNLDYAEDTHGNATIYWYEDETNNVDKLGDDTGTSYVRGGYLK
                     EIRYGQRAGALFSGTPQASHRIVLGHDERCLASGTGCDALTEDTRDNWPDVPFDRICK
                     DGDKCPYLDSPSFFTRKMLTSVSTYAWDAAAAATPAFSPVDTWTLEHLYLDPGDTGDST
                     DQSLWLDEIKHTGKRGTDLSLDAVKFSHVFMPNRVDGPSDDILSFERPRLRTVVSETG
                     AQTIVDYLPADCVAGQTMPRADENKRCFVYWSPYGQEEPILDWFQKYPVSSVRTTD
                     PLGGSEAVCHTYEYSGGGAWHYNDDPLTPAKERTWSQWRGYGKVTHYTGPSGGTRAKT
                     VTVYLRGMDGDRVLGDGKTPDPDRRRKAEVSGVTAAAVTDSDQLAGFQRESVAYDGD
                     KEVSGTVNDPWSKRTATQHKSYADTEAYYRVRGASHARTRITSRLNPYDRVRTTRIGY
                     DDYGMPVSVEDLGDDSVTGDEKCTRTWYARNDAAGLTALVSRTRIVGRACSVTDSALD
                     LPADAARPGDVVSDKATAYDTTWSATOKPTKGDARWSGRAKGYVGGDDRPLWQTAVT
                     DYDTLGRPVQVRNTDDVITSKTEYQPPAAGPLTSTVVNAKGHRTTVKDFALGADLK
                     VTDANGKVTESAYDSLGRVTSVMLPNRSRALGKTANYVYGYSVKSTSLPWVSSATLNG
                     DGSGYRTYEIYDSLLRTRQVOAPSAQQGRVIAQTLYDGRGLPVTAQADIWDDTAAPA
                     GKIVQIDGQAPRQTDSVYDGMGRVTKAVVKSYGVTEWAVDTAYRGDLVLTGAPEGGS
                     ANAVVTDAFGRTVERRDYAGTQPAGTDYMTTRYAFDAADRQKSITAHDRSAWTYTYDL
                     FGRQVSVTDPDKGTVTEYDALDRAVKSTDGRGEVLLFEYDVLGRKTGMWQSAKTGAN
                     KLAAWSFDTLAKGQQDTAVRYEGGETGRAYTQKVTRYDPLYKVTNNELTLPANDPLVA
                     AGYPARLAFSTGYNLDGTVKQAAPAVAGLSAETVSYTYDGLGQVLTAKGTTGYLQQA
                     AYSPLGDLRQMTLATDPTGAKKVYLNDVYEAGTRRLTRSVTDVHGFMLQELKYQQD
                     DAGNITSVSDATTLGGTGKGKADHQCFTYDGHRRLSEAWTPETADCSTSGRTVAGLGGAA
                     PYWTSYQIDDSGLRSKQTEHRMSGDDVTEYEYGTAEGQPHALSATVTGAENASYTYD
                     ETGNTETRPGVRATQTLDWNAEGRLAGVSEPAGGKPATGTAYYYDAGGDLLIRRPTT
                     TDGETVLILGTTEVHLKVSGNGAAKALSGARTYKAGSAVIAVRTSTAGVSGTKLITFLA
                     GDHHGTSGLAINADTLAFAKRWSTPFGAPRGTASGAWPDDRGLGKPADAATGLTQLG
                     ARQYPDTGRFLSVDPLLEPDKPNTLNGYAYASNSPVYNSDPSGTSGGWGGTAPTSSGGWTQPLTKQWTPGATYNF
                     ITRKSWDLPFNPPSQSLEEMLASMPDWGIVSDPKAANRWETSRSLPFGWLWGGGYPLRE
                     HQDPRGGDAFTSILAQDETISGLRSKMVGQARKGTAEKVEVFQYVDEGPEPGSPW
                     YKFNSLRGAAKDIAGVLITNGAVGTGNQADAFLGTYSGKARISSINKKEGSVTLKFSAW
                     NGSDWRSATHVVPRSWNPAFEDTFGAAVREDFSWEEKWPINECVNYSEWLE"
     stem_loop       8536..8571
                     /note="hairpin loop with 17/17 bp stem"
     gene            8602..9063
                     /gene="SC2H4.03"
     CDS             8602..9063
                     /gene="SC2H4.03"
                     /note="SC2H4.03, possible lipoprotein, len: 153 aa;
                     contains possible N-terminal signal sequence with
                     appropriately positioned PS00013 Prokaryotic membrane
                     lipoprotein lipid attachment site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="CAA20597.1"
                     /db_xref="PID:e1319221"
                     /db_xref="PID:g3559959"
                     /db_xref="GI:3559959"
                     /db_xref="SPTREMBL:086586"
                     /translation="MLKTPFAPAVRWTAAVLITLLAATACDDTGARAATVPADRVTGA
                     WQAGGGTALTLSADRTFTASGLDSGKLAGTGCPGGDTGGDWAFFADDGDGLYMTSPRA
                     RSGSEIGLSFAEQSDEPCRLTLTVVDDGDTLCATSDPETPCDLGVRFTRKD"
     misc_feature    8647..8679
                     /gene="SC2H4.03"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
```

```
LOCUS       SAAJ5198                 3240 bp    DNA             BCT       04-SEP-1998
DEFINITION  Streptomyces alboniger stgA, stgU and stgR genes.
ACCESSION   AJ005198
NID         g3550621
VERSION     AJ005198.1  GI:3550621
KEYWORDS    LysR-type transcriptional regulator; stgA gene; stgR gene; stgU
            gene.
SOURCE      Streptomyces anulatus.
  ORGANISM  Streptomyces anulatus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 3240)
  AUTHORS   Jimenez,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1998) Jimenez A., Universidad Autonoma de Madrid,
            Facultad de Ciencias, Centro de Biologia Molecular 'Severo Ochoa'
            (CSIC/UAM), Cantoblanco. Madrid, 28049, SPAIN
REFERENCE   2  (bases 1 to 3240)
  AUTHORS   Tercero,J.A., Espinosa,J.C. and Jimenez,A.
  TITLE     StgR, a new Streptomyces alboniger member of the LysR family of
            transcriptional regulators
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1..3240
                     /organism="Streptomyces anulatus"
                     /db_xref="taxon:1892"
     gene            complement(180..1388)
                     /gene="stgA"
     CDS             complement(180..1388)
                     /gene="stgA"
                     /function="putative aminotransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="StgA protein"
                     /protein_id="CAA06435.1"
                     /db_xref="PID:e1318407"
                     /db_xref="PID:g3550622"
                     /db_xref="GI:3550622"
                     /db_xref="SPTREMBL:O88124"
                     /translation="MEFRQSSKLSEVCYEIRGPVIEHANALEEAGHSVLRLNTGNPAL
                     FGFEAPEEIVQDMIRMLPQAHGYTDSRGVLSARRAVAQRYQALGLDEVSVDDVFLGNG
                     VSELVTMAVQALLEDGDELIPAPDFPLWTAVITLSGGKAVHVVCDEAADWYPDLDDM
                     ASKITDQTKAVEIINPTGAVYPKEILEGILDLARRHGLMVFADEIYDQILYDDAVHHS
                     VAALAPDLVVLITFCGLSKTYRVAGFRSGWLVITGPRQHARDYLEGLTMLASMRLCPNA
                     PAQFAIQAALGRGRGSIRELTAPGGALHEQRDRAWEKLNEIPGVSCVKPGALYAFPRL
                     DPKVHPIHDDEKFVLDLLLQEKIQVVQGTGFNWPRPDHFRILTLPHADDLDAAISRIG
                     RFLSGYRQGR"
     gene            1473..2105
                     /gene="stgU"
     CDS             1473..2105
                     /gene="stgU"
                     /codon_start=1
                     /transl_table=11
                     /product="StgU protein"
                     /protein_id="CAA06436.1"
                     /db_xref="PID:e1318408"
                     /db_xref="PID:g3550623"
                     /db_xref="GI:3550623"
                     /db_xref="SPTREMBL:O88133"
                     /translation="MSPRRSYDQYCAAARALDAVGDRWTLLVVRELLVGPRRYTDLHA
                     DLPGVSTDVLASRLKDMEQQGLTTRRRLPPPAAAYVYELTLKGRQLLPVLEALAAWGV
                     HGLEERRPIDAVRAHWFALPMLRRLARIGAEGTVEVGLGEGRFHVRLGDGAETYGDGP
                     APDAPEVRLALDAQTCVAVSRGELTVGEGVRAGRIELSGEGTLAKALLAT"
     gene            2219..3202
                     /gene="stgR"
     CDS             2219..3202
                     /gene="stgR"
                     /codon_start=1
                     /transl_table=11
                     /note="LysR-type transcriptional regulator"
                     /product="StgR protein"
                     /protein_id="CAA06437.1"
```

```
     gene            attachment site"
                     complement(9074..10282)
                     /gene="SC2H4.04c"
     CDS             complement(9074..10282)
                     /gene="SC2H4.04c"
                     /note="SC2H4.04c, probable aminotransferase, len: 402 aa;
                     smilar to many e.g. AAT_BACST aspartate aminotransferase
                     (EC 2.6.1.1) (393 aa), fasta scores; opt: 657 z-score:
                     607.8 E(): 1.3e-26, 33.9% identity in 386 aa overlap.
                     Contains Pfam match to entry PF00155 aminotran_1,
                     Aminotransferases class-I, score 136.00, E-value 6.9e-37"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aminotransferase"
                     /protein_id="CAA20598.1"
                     /db_xref="PID:e1319222"
                     /db_xref="PID:g3559960"
                     /db_xref="GI:3559960"
```

```
Query Match             6.5%;  Score 107.2;  DB 1;  Length 25970;
Best Local Similarity  50.8%;  Pred. No. 3e-07;
Matches  365;  Conservative    0;  Mismatches  338;  Indels   15;  Gaps    4;

Qy    268  cgcgagcgtggagagagggagcgggccgcgcccgtgctgccgccaccggtgaccgtc   327
           ||||| || ||| ||   || ||  |  || || ||||| |||| |||| || ||||
Db  10214  CGCCAACGCGCTGGAGGAGGCGGGGGCACAGCGTGCTGCCCTGAACACCGGCAACCCCC  10155

Qy    328  cgtgttcccggccttcccgcacggccgtcgaggcgaagacgccgtcgcgccgcgctgcg   387
           |  |||||   |||||||||  ||  || || || ||  || || || |||| ||||||
Db  10154  CCTCTTC---GGGTTCGAGGCGCCCGGAGGAGATCGTCCAGGACACGATCCGCATGCTGCC  10098

Qy    388  caccggccagttcaactgctacgcgcgcggtcggcctcccgccgcacgaagcgccct   447
           |  || ||||  || ||| ||  || ||||| |||| |||| |||| |||| || ||
Db  10097  GAGGGCGC------ACGGCTACACGGACTTCGCGCGGCATCCTCTCCGCCCCGGGCCGT  10044

Qy    448  agcagagcacttgtcacagggcgtgcctacagctatcggccgacgacgtcttcctcac   507
           |||| ||||  |  |  | ||| ||  |||||| |  || ||| ||| |||||||| |
Db  10043  CGCCCAGCGC---TACCAGCGCGCTCGGCCTGGAGTGGACGTGGACGAGCGTCTTCCTGGG   9987

Qy    508  cgcgcggcggaactcaggcgtcgatcgaagtcataatcccggtgctggccccagactgccgggcgc   567
           ||| ||| || |  ||||  || || ||    || |||||  || || ||||| | |   |||  |
Db   9986  CAACGGCGGTCTCGGAGCGTGATCTCCATGGCCGTGCAGGCCCTCCTTGAG---GACGGCGA   9930

Qy    568  caacatactgcttcccggccaggctatccaaattacgaggcgcagccgcattcaacca   627
           |||| |  || |||| ||   || |||||| || || || ||| ||  | || || ||
Db   9929  CGAGGTGCTGATCCCGCCCCGACTTCCCCTCTGGACGGGCGGTGACCACCCTCGCGGG   9870

Qy    628  gctggagtccggcacttgacctcatcccgacaagggtgggagatcgacatcgactc   687
           |||  || | ||||   | ||| |    || || ||   |||| ||| ||| || |
Db   9869  CGGCCAAGGCGGTGCACTACCTCTGCGACGAGCAGGCCGAGTGGTACCCCGACCTGGCCGA   9810

Qy    688  gctggaatccatcgccgacaagaacaccgcgatggtcatcataaacccaacaatc   747
           |||  |  ||||| || ||||||  |||| | ||  |||| || |||| ||| ||
Db   9809  CATGGAGGCGGAAGATCACCGCAACCCAAGGCGTCGTTCATCATCAACCCCACAACCC   9750

Qy    748  gtgcgcagcgtttactcctacgacatctggccaggtgcgcgaggtgcaaggaagct   807
           |||| || || ||| || || || |||||  |||||| || || || ||| |||
Db   9749  CACCCGGCGCCGGTCTACCCCAAGGAGATCGTCGAGGGCATCCTCGACCTCGCCCGGCA   9690

Qy    808  cggaatattggtgatcggctgacgaggtttacggcgaaactggttctggcagcgccccgtt   867
           |||  |  ||||| |||| | ||||| ||||||| ||  |||| || || |||| |||
Db   9689  CGGGCTGATGGTCCTCGCCGACGAGATCTACGACCAGATCCTGTACGACGACGCCGTGCA   9630

Qy    868  tatccgatcgggctcttgggcacattgcccggtcttgtccattggatctctgtccaa   927
           || ||  |||| || || ||||| ||  || ||| || ||||||  || | ||||||
Db   9629  CCACTCGGCCGGCGTCCCTGCCCCCGACCTGGTGGTGCTGACCTTCTGCGGCCTGTCCAA   9570

Qy    928  gtcgtggatagtgcctggatggcggcgacttgatggtggccgtgtacgaccccacaaag   985
           | || |  |  || || || || ||||  | |||| ||| ||| |||| |||  ||| |
Db   9569  GACGTACCGGGTGGCCGGGTTCCGCTCCGGCTGGCTGGTCATCACCGGGCCGCGCCAGCAG   9512
```

RESULT .12

```
                    /db_xref="PID:e1318409"
                    /db_xref="PID:g3550624"
                    /db_xref="GI:3550624"
                    /db_xref="SPTREMBL:O88132"
                    /translation="MDLDLALVRSFAITAEELHFRRAAERLFITQQGLSQRIRRLELL
                    LGVSLFVRGNRSVALTEEGQRFLAPARRVMEAADAAVQVPRERRPLRVDVVDGRLA
                    PLQIVRKLLEEDPELPLNLSMRQSLAAALPALARDEIDTAFGRVHDVQGWSAEFSHRL
                    VRLEPLRALVPRTHPLAGRDEVRIDELRSCAIWQPRNGVPSEWEHYVGRLADSFGLSL
                    AFSGAALSDEHFLEHLWSHGEMACLAGADVSYAFSPDIRSIPLVDPTPVYPWSMVWRR
                    QAGHPLVDRLVGLAKRSAGDWLTHVPGEIWLPAPDRALLRGAGVDVDAVRG"
BASE COUNT     447 a    1112 c    1189 g    492 t
ORIGIN
```

```
Query Match           6.3%;    Score 104.8;    DB 1;    Length 3240;
Best Local Similarity 51.4%;   Pred. No. 8.4e-07;
Matches  300;  Conservative    0;  Mismatches  272;  Indels   12;  Gaps    2;

Qy     395 cagttcaactgctacggccgccggctcggcctcccccggccgcacgaagcggcctagcagag   454
           ||| |||||| ||| || ||||| || ||   | |  |||||  || ||| ||||||||||
Db    1202 CAGGGCGCACGGCTACACGGACTCGCGCGGGGTCCGTCCGGCCGCGCGGCGCCGTCGCCCAG  1143

Qy     455 cacttgtcacaggggcgtgcgcctccaagctatcggccgacgagctcttcctcaccgcggc   514
           | ||  ||| || || || ||||| ||||   |||| ||| ||||  |||||||||| |||
Db    1142 CGCTATCAGGGCGGCTCGGCCTCGACGACGAGGTCTCGGTGGACGACGTCTTCCTCGGCAACGGC  1083

Qy     515 ggaactcaggcgatcgaagtcataatcccggtgctggcccagactgccggcggccaacata   574
           |  | |||| | | || ||  ||||  || ||| | ||| |  |||||  |||| | ||
Db    1082 GTGTCCGAGCTGGTCACGATGGCGGTCCAGGCGCT---CCTGGAGGACGGGCGACGAAATC  1026

Qy     575 ctgcttccccgccaggctatccaaattacgagccgcgagcggcattcaacaagctggag   634
           ||| || ||||  || | |  | |||| ||| |  |||| ||||| | |||||  ||||
Db    1025 CTCATCCCCGCGCCCGACTTCCCGCTGTGGACCCGCGGTGACGACCCTCCGGAGGCAAG   966

Qy     635 gtccggcacttcgacctcatcccgacaaggggtgggagatcgacatcgcgctggaa   694
           | |||| ||| || ||||| ||  ||||| ||| |||| ||| || ||| ||||
Db     965 GCCGTCCACTACGTCTGCGACGAGGCGCCGGACCTGGTACCCCGACCTCGACGACATGGCC   906

Qy     695 tccatcgccgacaagaacacaccgcgatgtcatcataaacccaaacaatccgtgcggc   754
           | ||| ||| ||| || |||| ||||  || | | ||| ||||| ||||| |||||
Db     905 TCGAAGATCACCGACCAGGACACCAAGGCCGTCGAGATCATCAACCCCA-------CGGGC   855

Qy     755 agcgttactcctacgacccatctggccaaggtcggcgaggtggcaaggaagctcggaata   814
           ||||  |||  |||| ||  ||| |||||| ||||||| ||  |||| |||| ||  ||
Db     854 GCGGTCTATCCGAAGGAGATCCTCGGAGGGCATCCGATCTCGCGGCGGCGGCGCACGGCCTG   795

Qy     815 ttggtatcgctgacgaggtttacggcaaactggttctgggcagcgcccgtttatcccg   874
           || |||| |  |||||   | ||| |||| ||| ||| |||| |||| | ||||| ||
Db     794 ATGGTCTTCGCCGACGAGATCTACGACGACCAGATCCGTGTACGACGACGACCCGTCACCACAGC   735

Qy     875 atgggcgtcttggcacattgccccgtgtcttgtccattggatcctgtccaagtcgtgg   934
           ||||| ||| | ||  |   |||  || |||| |  | |||| |||| ||||| ||
Db     734 GTGGCGCCCCTCCCCGGATCTCGTGTCCTCACCTTCTCGCGGTCTTTCGAAGACGTAC   675

Qy     935 atagtgcctgatggcggcgactggaggtggcggtgtacgaccg   978
           ||  |||||| | |||  || | |  ||| ||| ||  |||
Db     674 CGCGTGGCGGGTTCCGTTCGGGCTGGCTGGTCATCACCGGCCC   631
```

```
RESULT 13
SAN6517/c
LOCUS      SAN6517      3240 bp       DNA            BCT       20-OCT-1998
DEFINITION Streptomyces anulatus stgA, stgU and stgR genes.
ACCESSION  AJ006517
NID        g3550597
VERSION    AJ006517.1  GI:3550597
KEYWORDS   aminotransferase; LysR-type transcriptional regulator; putative;
           stgA gene; stgR gene; stgU gene.
SOURCE     Streptomyces anulatus.
  ORGANISM Streptomyces anulatus
           Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
           Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE  1 (bases 1 to 3240)
```

AUTHORS   Jimenez,A.
TITLE   Direct Submission
JOURNAL   Submitted (01-JUN-1998) Jimenez A., Universidad Autonoma de Madrid,
           Facultad de Ciencias, Centro de Biologia Molecular 'Severo Ochoa'
           (CSIC/UAM), Cantoblanco. Madrid, 28049, SPAIN
REFERENCE 2 (bases 1 to 3240)
AUTHORS   Tercero,J.A., Espinosa,J.C. and Jimenez,A.
TITLE   StgR, a new Streptomyces alboniger member of the LysR family of
           transcriptional regulators
JOURNAL   Mol. Gen. Genet. 259 (5), 475-483 (1998)
MEDLINE   99005239

```
FEATURES             Location/Qualifiers
     source          1..3240
                     /organism="Streptomyces anulatus"
                     /db_xref="taxon:1892"
     gene            complement(180..1388)
                     /gene="stgA"
     CDS             complement(180..1388)
                     /gene="stgA"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aminotransferase"
                     /protein_id="CAA07075.1"
                     /db_xref="PID:e1318411"
                     /db_xref="PID:g3550598"
                     /db_xref="GI:3550598"
                     /db_xref="SPTREMBL:O88124"
                     /translation="MEFRQSSKLSEVCYEIRGPVIEHANALEEAGHSVLRLNTGNPAL
                     FGFEAPEIVQDMIRMLPQAHGYTDSRGVLSARRAVAQRYQALGLDEVSVDDVFLGNG
                     VSELVTMAVQALLEDGDEILIPAPDFPLWTAVTTLSGKAVHVCDEAADWYPDLDDM
                     ASKITDQTKAVEIINPTGAVYPKEILEGILDLARRHGLMVFADEIYDQILYDDAVHHS
                     VAALAPDLVLLTFCGLSKTYRVAGFRSGWLVTGPRQHARDYLEGLTMLASMRLCPNA
                     PAQFAIQAALGRQSIRELTAPGGALHEQRDRAWEKLNEIPGVSCVKPKGALYAFPRL
                     DPKVHPIHDDEKFVLDLLQEKIQVVQGTGFNWPRPDHFRILTLPHADDLDAAISRIG
                     RFLSGYRQGR"
     gene            1473..2105
                     /gene="stgU"
     CDS             1473..2105
                     /gene="stgU"
                     /codon_start=1
                     /transl_table=11
                     /product="StgU"
                     /protein_id="CAA07076.1"
                     /db_xref="PID:e1318412"
                     /db_xref="PID:g3550599"
                     /db_xref="GI:3550599"
                     /db_xref="SPTREMBL:O88133"
                     /translation="MSPRRSYDQYCAAARALDAVGDRWTLLVVRELLVGPRRYTDLHA
                     DLPGVSTDVLASRLKDMEQQGLTTRRRLPPPAAYYVELTLKGRQLLPVLEALAAWGV
                     HGLEERRPTDAVRAHWFALPMLRTLARIGAEGTVEVGLGEGRFHVRLGDGAETYGDGP
                     APDAPEVRLALDAQTCVAVSRGELTVGEGVRAGRIELSGEGTLAKALLAT"
     gene            2219..3202
                     /gene="stgR"
     CDS             2219..3202
                     /gene="stgR"
                     /codon_start=1
                     /transl_table=11
                     /product="LysR-type transcriptional regulator"
                     /protein_id="CAA07077.1"
                     /db_xref="PID:e1318413"
                     /db_xref="PID:g3550600"
                     /db_xref="GI:3550600"
                     /db_xref="SPTREMBL:O88132"
                     /translation="MDLDLALVRSFAITAEELHFRRAAERLFITQQGLSQRIRRLELL
                     LGVSLFVRGNRSVALTEEGQRFLAPARRVMEAADAAVQVPRERRPLRVDVVDGRLA
                     PLQIVRKLLEEDPELPLNLSMRQSLAAALPALARDEIDTAFGRVHDVQGWSAEFSHRL
                     VRLEPLRALVPRTHPLAGRDEVRIDELRSCAIWQPRNGVPSEWEHYVGRLADSFGLSL
                     AFSGAALSDEHFLEHLWSHGEMACLAGADVSYAFSPDIRSIPLVDPTPVYPWSMVWRR
                     QAGHPLVDRLVGLAKRSAGDWLTHVPGEIWLPAPDRALLRGAGVDVDAVRG"
BASE COUNT     447 a    1112 c    1189 g    492 t
ORIGIN
```

JOURNAL   Unpublished (1996)
COMMENT   This sequence was generated by the Genome Sequencing Center at
          Genome Therapeutics Corporation (Collaborative Research Division),
          100 Beaver St., Waltham, MA, 02154. Please contact Doug Smith
          (smith@cric.com) for further information. The sequence represents
          the insert of a Lorist 6 cosmid clone from a mapped set of clones
          constructed from M. leprae genomic DNA isolated from armadillo
          liver [3]. The sequence may not represent the entire cloned insert of
          the cosmid if an overlapping region was previously sequenced from
          another clone. Coding sequences larger than 60 amino acids were predicted
          on the basis of codon usage and homology information. An attempt was
          made to locate the most probable start site based on codon usage,
          homology, the presence of a Shine-Dalgarno sequence, or overlapping orf that
          suggested translational coupling. It is possible that the actual
          start site differs from the one selected.

FEATURES   Location/Qualifiers
     source   1..39399
                   /organism="Mycobacterium leprae"
                   /specific_host="Dasypus novemcinctus"
                   /db_xref="taxon:1769"
                   /clone="cosmid B1970"
                   /tissue_lib="Lorist 6"
                   /note="The liver of the armadillo was used to isolate the
                   Mycobacterium leprae."
BASE COUNT   7827 a 10529 c 12083 g   8960 t
ORIGIN

Query Match   5.7%; Score 95; DB 1; Length 39399;
Best Local Similarity 50.1%; Pred. No. 1.7e-05;
Matches 290; Conservative 0; Mismatches 285; Indels 4; Gaps 2;

Qy   405 gctacggcgcggcgtggcctcccgccccgcacgaagcgcgtagcagagcacttgtcac 464
           ||| |||||| || ||| ||||||||||| ||| |||| | ||||||||||
Db   9552 GGTACTCGACTCCCAGGGCATCCTGCTGCTCGACGTCGGCGGTGGTTACCCGTACCAGAGC 9493

Qy   465 agggcgtgcctacaagctat---cggccgacgacgtcttcctcaccgccggcggactc 521
           ||| |||||||||||||| ||||||||||||||||||||| || | |||||
Db   9492 TGGTCGACGGGTTCCCCGATTCGACGTGCGATGAGCGTCTATCTCGGCAACGGCGGTCTCCG 9433

Qy   522 aggggatcgaagtcatatcccggtgctgccagactgccgcgccaacatactgttc 581
           || ||||||| ||||| |||||| |||||||| |||| | ||||||| |
Db   9432 AGCTGATCACGATGACGCCTGCAGGCCCGTTAGACACACGGC-GATCCAAGGTGCTGATCC 9374

Qy   582 ccggccaggtcatccaaattacgaggcgcggcgagcggcattcaacaagctgagtcggc 641
           ||||||||| | |||| ||||||| ||||| |||| |||||||| |||
Db   9373 CCTGCGGACGTACCCGTGTGGACAGCGTCGACCTCGCTGGCGGGACACCCGTAC 9314

Qy   642 acttcgacctcatcccgacaaggggtgggagatcgacatcgactcgctggaatccatcg 701
           || ||| ||||| |||||||||| |||| ||||||| ||||||| || ||||||| |
Db   9313 ATTACCTCTGCGACGAAACCCAAGGCTGCAACCAGATATCGCCGACGTGGAATCTAAGA 9254

Qy   702 ccgacaagaaccaccacgggatggtcatcataaaccccaaacaatccgtgcggcagcgttt 761
           || || |||| ||||||| ||||||||||||||||||||||| | |||| |||| |||||
Db   9253 TCACCGAGCGCCACCAAAGCACTGGTCGTCGTGATCAACCCCAACAACCCTACCGGCGCCTCT 9194

Qy   762 actcctacgaccatggccaaggtcgcgaggtggcaaggaagctcggaatattggtga 821
           |||| || |||||||||||||| ||||||||| ||||| || |||| ||||||
Db   9193 ACAGCAACGAAAATCCTTAACCAAATAGTCGACTTAGCACCTAGCATCAGCTGCTGC 9134

Qy   822 tcgctgacgaggtttacggcaaactggttctggccagcgccgttatccgatgggcg 881
           ||| ||||||||||| || |||||| |||||||||| ||||| |||||||||||
Db   9133 TTGCCGACGAGATCTACGACAAAGAATTCTCATGACGACCAAAGATTCTCATGACGACCACCAGCGTGGCCT 9074

Query Match   6.3%; Score 104.8; DB 1; Length 3240;
Best Local Similarity 51.4%; Pred. No. 8.4e-07;
Matches 300; Conservative 0; Mismatches 272; Indels 12; Gaps 2;

Qy   395 cagttcaactgctacggcgcggcgtggcctcccgcacgagagcgcgtagcagag 454
           ||| ||||||| || ||| ||||||||||| |||| |||||| | ||||||||||
Db   1202 CAGGGCCACGGCTACACGGACTCGCGGCGTTCCGTCCGCCGCCGCCGTCGCCCAG 1143

Qy   455 cacttgtcacagggggtgcctacaagctatcggccgacgacgttcctcaccgccggc 514
           || |||||| ||||| ||||||||||||||| |||||||||| |||||||||| |||
Db   1142 CGCTATCAGGCGCTCGGCCTCGACGAGGTCTCGGTGGACGAGCGTCTTCCTCGGCAACGGC 1083

Qy   515 ggaactcaggcgatcgaagtcataatcccggtgctggcccagactgccggcgccaacata 574
           || |||||| ||||| | ||| ||||| ||||||||| || |||| |||| || ||||
Db   1082 GTGTCCGAGCTGGTCACGATGGCGGGTCCAGGCGCT---CCTGGAGGACGGCGACGAAATC 1026

Qy   575 ctgcttcccgcgcaggctatccaaattacgaggcgcgagcggcattcaacaagctggag 634
           ||| |||||| || ||||||| |||||||| ||| || ||||||| |||| || |
Db   1025 CTCATCCCCGGCGCCCGACTTCCGCTGGACCGCGTGAGGACCCCTTCCGGAGGCCAAG 966

Qy   635 gtccggcacttcgacctcatcccgacaagggggtgggagatcgacatcgactcgctggaa 694
           ||||| |||||||| ||||| |||||||| |||| ||||||||| |||||| |||
Db   965 GCCGTCCACTACGTCTCGCGACGAGGGCGGCGGACTGGTACCCCGACCTGACGACATGGCC 906

Qy   695 tccatcgccgacaagaacacaccgatggtcatcataaacccaaacatccgtgcggc 754
           ||| ||||||||||| || || ||||||||||||||||| | |||| |||| |||
Db   905 TCGAAGATCACCGACCAGACCAAGGCCGTCGAGATCATCAACCCA--------CGGGC 855

Qy   755 agcgttactctacgaccatctggccaaggtcgcggaggtgcaaggaagctcggaata 814
           || ||| ||||| || || |||||||||| ||||||||| ||||| || |||| ||
Db   854 GCGGTCTATCCGAAGGAGATGATCCTGGAGGGCATCCTCGATCTCGCGGCGGCACGGCCTG 795

Qy   815 ttggtgatcgctgacgaggtttacgggcaaactggttctggcagcgccccgtttatcccg 874
           ||| | |||| ||||||| || || ||||| |||| |||| |||||| ||| |
Db   794 ATGGTCTTCGCCGACGAGATCTACGACCAGATCCTACGACGACGCCGTGCCACCACAGC 735

Qy   875 atgggcgtctttgggcacattgccccggtcttgtccattggatctctgtccaagtcgtgg 934
           ||||| | |||| ||| || |||||||| |||| ||||| |||||| ||||| ||||||
Db   734 GTGGGCCGCCCTCGCCCCGGATCTCGTGTCCTCCACCTTGCGGTCTTGCGAAGACGTAC 675

Qy   935 atagtgcctggatggcgcacttggatggtggtggttgtgacc 978
           |||||||| | ||| | || || |||||| ||||||
Db   674 CGCGTGCGCGGGTTTCCGTTCGGCGTCGGCTGCTGGTCATCACCGGCCC 631

RESULT 14
MSGB1970CS/c

LOCUS   MSGB1970CS   39399 bp   DNA   BCT   26-JUN-1996
DEFINITION Mycobacterium leprae cosmid B1970 DNA sequence.
ACCESSION   L78815
NID   g1377775
VERSION   L78815.1 GI:1377775
KEYWORDS   .
SOURCE   Mycobacterium leprae (clone: cosmid B1970) (tissue library: Lorist
            6) DNA.
  ORGANISM Mycobacterium leprae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
REFERENCE   1 (sites)
  AUTHORS   Eiglmeier,K., Honore,N., Woods,S.A., Caudron,B. and Cole,S.T.
  TITLE   Use of an ordered cosmid library to deduce the genomic organization
            of Mycobacterium leprae
  JOURNAL   Mol. Microbiol. 7 (2), 197-206 (1993)
  MEDLINE   93188700
REFERENCE   2 (bases 1 to 39399)
  AUTHORS   Smith,D.R., Richterich,P., Rubenfield,M., Butler,C., Lee,H.-M.,
            Xu,Q., Gunderson,K., Chung,M., Maher,J.K., Deloughery,C.,
            Aldrich,T., Imrich,J., Tulig,C., Smyth,A., Drill,S., Avruch,A.S.,
            Rice,P., Abendschan,K., Aldredge,T., DeLoughery,C., Kirst,S.,
            Safer,H., Connelly,C., McDougall,S., Eiglmeier,K., Bergh,S.,
            Cole,S., Robison,K., Jaehn,L., Gryan,G., Church,G.M. and Mao,J.
  TITLE   Prepublication submission

```
Qy    882  tcttgggcacattgccccggtcttgtccattggatctctgtccaagtcgtggatagtgc  941
           ||| ||  ||||||| || |||||| || |  |||||  ||| ||| |||||| |||| ||
Db   9073  CAATTGCGCCCGATCTACTGTGCTTGACCTTCAACGGTCTTTCGAAAGCCTACCGTGTCG  9014

Qy    942  ctggatggcgacttggatgggtgcggtgtacgacccca  980
           |-||| ||||||  | ||||||-|-| |||||- || ||||
Db   9013  CCGGGTACCGGTCCGGCTGGCTAGCGATCACCGGGCCCA  8975
```

RESULT 15
ATF9D16
LOCUS   ATF9D16   119430 bp   DNA   PLN   25-MAR-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F9D16 (ESSAII
            project).
ACCESSION   AL035394
NID   g4454022
VERSION   AL035394.1   GI:4454022
KEYWORDS   .
SOURCE   thale cress.
ORGANISM   Arabidopsis thaliana
            Eukaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
            Rosidae; Capparales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 119430)
AUTHORS   Bevan,M., Wedler,H., Wedler,E., Wambutt,R., Hoheisel,J.,
            Mewes,H.W., Mayer,K.F.X. and Schueller,C.
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 119430)
AUTHORS   EU Arabidopsis sequencing project.
TITLE   Direct Submission
JOURNAL   Submitted (09-FEB-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT   Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosome 4 can be
            viewed at: http://websvr.mips.biochem.mpg.de/proj/thal/.
FEATURES   Location/Qualifiers
     source   1..119430
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene   1..4213
                     /gene="F9D16.10"
     CDS   complement(join(>1..133,216..272,344..437,629..810,
                     1176..1403,1480..1661,1900..2110,2438..2529,2696..2953,
                     3070..3322,3371..3616,3704..4213))
                     /gene="F9D16.10"
                     /note="similarity to hypothetical proteins from human,
                     Arabidopsis, yeast"
                     /codon_start=1
                     /product="putative protein (fragment)"
                     /protein_id="CAA23020.1"
                     /db_xref="PID:e1386693"
                     /db_xref="PID:g4454023"
                     /db_xref="GI:4454023"
                     /translation="MAAFEDRADENDEIAFKDGNTDISQQLMDRYGKSSAAQHRHLVA
                     TAVAMRSILTSESLPSPSAFFAAAISSVDSSTEDPVAVSALLTFLSIVPLVPSGEI
                     LKFAIDKRPKVRRCAQECLEKLFGSLRSSTVIKEASNTVALLKEYKPVLSDLSSTKI
                     EEGSKVDSTLKSENAEAAHVLNVLSATIPFLSAKFSPLRQILKGIDTIFKNSEDKIV
                     VPEIEGLVTTLTSYLSLHDKNPADTVHVTLLKDALEKAYSVEPTLCLSKLPLVCSS
                     LAGLLTSTDDIASKASTILKDLISSHIDKKLLTEGSLSNQDEDNVTSGDNINAARCV
                     CSVFESVLNSCDGIPNEHILTVINLLIEKLGELSYILAKNIILKLADLMKNAIGDTSS
                     SQDLQOCIGSAVVAMGPVRLLTLLPITLHAESHSCTNAWLIPILRKYIIGASLAYVD
                     NIVPLAKSLMLASKGAKKSTHGKELRACGHELLRLLPAFCNYPVDVANKFGSLAKLMV
                     KFIKKSFMHEAVALSLQMLVNQNKGMPKPSTDMGEAKAISEEDATTELERGFHYSKK
                     ASTKNMKALASSSTELLQTLVDVFTVSGTEISADFKAAIGCLASTLDSSVRKKIIISL
                     LNKFDPAGESETEGQVNQSNDSVDEEKENCSSTKTQLKRSAVLDLASSFVEGAKEDLI
                     ELIIYNLVRQSFQATDEADLYGAYDTLSRVLQEHGWFCASHFAEVIEMLLSHKTPEDAA

```

Right column (feature table, continued):

```
exon   SSRSRFACLHVLMAHGI"
              complement(1..133)
              /gene="F9D16.10"
              /number=1
intron   complement(134..215)
              /gene="F9D16.10"
              /number=1
exon   complement(216..272)
              /gene="F9D16.10"
              /number=2
intron   complement(273..343)
              /gene="F9D16.10"
              /number=2
exon   complement(344..437)
              /gene="F9D16.10"
              /number=3
intron   complement(438..628)
              /gene="F9D16.10"
              /number=3
exon   complement(629..810)
              /gene="F9D16.10"
              /number=4
intron   complement(811..1175)
              /gene="F9D16.10"
              /number=4
exon   complement(1176..1403)
              /gene="F9D16.10"
              /number=5
intron   complement(1404..1479)
              /gene="F9D16.10"
              /number=5
exon   complement(1480..1661)
              /gene="F9D16.10"
              /number=6
intron   complement(1662..1899)
              /gene="F9D16.10"
              /number=6
exon   complement(1900..2110)
              /gene="F9D16.10"
              /number=7
intron   complement(2111..2437)
              /gene="F9D16.10"
              /number=7
exon   complement(2438..2529)
              /gene="F9D16.10"
              /number=8
intron   complement(2530..2695)
              /gene="F9D16.10"
              /number=8
exon   complement(2696..2953)
              /gene="F9D16.10"
              /number=9
intron   complement(2954..3069)
              /gene="F9D16.10"
              /number=9
exon   complement(3070..3322)
              /gene="F9D16.10"
              /number=10
intron   complement(3323..3370)
              /gene="F9D16.10"
              /number=10
exon   complement(3371..3616)
              /gene="F9D16.10"
              /number=11
intron   complement(3617..3703)
              /gene="F9D16.10"
              /number=11
exon   complement(3704..4213)
              /gene="F9D16.10"
              /number=12
exon   9624..9967
              /gene="F9D16.20"
              /number=1
```

```
Qy    500 ttcctcaccgccggcggaactcaggcggatcgaagtcataatcccggtgctgcccagact 559
           || ||| ||| |||| |||| || || |||| | || ||||| |  ||||||| |||||||
Db  29169 TTTATGACTCTGGGATGCAAACAAGCTATTGAGCTCGCGGTAGACATTCTCGCTAAACCG 29228

Qy    560 gccggcgccaacatactgcttcccgccaggctatccaaattacgaggcgcgagcggca 619
           |  ||||| |||| | ||||||| ||| ||| | ||||||||||| ||| ||||||||
Db  29229 AAAG---CCAACGTTTGCTTCCGAGTCCCGGCTTCCCATGGGACCTAGTCCGCTCCATC 29285

Qy    620 ttcaacaagctggagtccggcacttcgacctcatcccgacaagggtgggagatcgac 679
           | ||| |||||||||| ||||| |||| || || ||| |||  || ||||| ||||||
Db  29286 TACAAGAACCTTGAGTCGCCGCCACTATAATTTCCTTCCAGAAAAGAACCTTGAAATCGAC 29345

Qy    680 atcgactcgctggaatccatcgccgacaagaacaccaccgcgtggtcatcataaaccca 739
           | ||| ||| |||||||||||| || |||||||| ||| ||| ||||| ||||
Db  29346 TTTGATAGCGTCCGAGCGTCGTGGACGAGAACACATTTGCCATATTTATAATCAACCCC 29405

Qy    740 aacaatccgtgcggcgagcgttactcctacgaccatctggccaagt 786
           ||| ||||| ||| | || |||| |||| | |||||||
Db  29406 CACAACCCCAATGGTAACACCTACTCCGAGGCTCATCTCAAACAGGT 29452


Search completed: October  1, 1999, 15:32:18
Job time: 6041 sec
```

```
gene          9624..10881
              /gene="F9D16.20"
CDS           join(9624..9967,10214..10330,10428..10881)
              /gene="F9D16.20"
              /note="similarity to AR411 - Arabidopsis thaliana
              PATX:G1669603"
              /codon_start=1
              /product="putative protein"
              /protein_id="CAA23021.1"
              /db_xref="PID:e1386934"
              /db_xref="PID:g4454024"
              /db_xref="GI:4454024"
              /translation="MDEGDLEAIVRGYSGSGDAFSGESSGTFSPSFCLPMETSSFYEP
              EMETSGLDELGELYKPFYPFSTQTILTSSVSLPEDSKPFRDDKKQRSHGCLLSNGSRA
              DHIRISESKSKKSKKNQQKRVVEQVKEENLLSDAWAWRKYGQKPIKGSPYPRSYYRCS
              SSKGCLARKQVERNPQNPEKFTITYNEHNELPTRRNSLAGSTRAKTSQPKPTLTKK
              SEKEVVSSPTSNPMIPSADESSVAVQEMSVAETSTHQAAGAIEGRRLSNGLPSDLMSG
              SGTFPSTGDFDELLNSQEFFSGYLWNY"
intron        9968..10213
              /gene="F9D16.20"
              /number=1
exon          10214..10330
              /gene="F9D16.20"
              /number=2
intron        10331..10427
              /gene="F9D16.20"
              /number=2
exon          10428..10881
              /gene="F9D16.20"
              /number=3
repeat_region 10899..10931
              /note="33bp ta tandem repeat"
exon          complement(11426..11752)
              /gene="F9D16.30"
              /number=1
gene          11426..13581
              /gene="F9D16.30"
CDS           complement(join(11426..11752,11828..11994,12098..12206,
              12305..12457,12570..12647,12732..12932,13067..13279,
              13390..13581))
              /gene="F9D16.30"
              /note="strong similarity to cellulase - Phaseolus
              vulgaris, PID:g1039431; contains EST gb:F14141, T76139,
              F13971"
              /codon_start=1
              /product="putative cellulase"
              /protein_id="CAA23022.1"
              /db_xref="PID:e1386935"
              /db_xref="PID:g4454025"
              /db_xref="GI:4454025"
              /translation="MGKLLVLMLVGMFLAFESLEALEYGDALNKSILFFEQRSGKLP
              TNQRVKWRADSALSDGSLANVNLIGGYDAGDNVKFWPMSFTTLLSWAAIEYQNEI
              SSVNQLGLRSTIKWGTDFILRAHTSPMLYTQVGDGNSDHSCWERPEDMDTSRTLYS
              ISSSPGSEAGEAAAALAAASLVFKSVDSTYSSTLNHAKTLFEFADKYRGSYQASC
              PFYCSYSGYQDELWAAAWLYKATGDKIYINVYISNKDWSQAVNEFSWDNKFVGAQAL
              LVSEFYNGANDLAKFKSDVESFVCAMMPGSSSSQQIKPTPGGLLFIRDSSNLQYVTAT
              TVLFHYSKTLTKAGVGSIQCGSTKFTVSQIRNFAKSQVDYILGNPMKMSYMVGFGTK
              YPTQPHHRGSSLPSIQSKPEKIDCNGGYSYYNSDTPNPNVHIGAIVGGPNSSDQYSDK
              KSDYSHAEPTTYINAAFIGPVAALISSSG"
intron        complement(11753..11827)
              /gene="F9D16.30"
              /number=1
exon          complement(11828..11994)
              /gene="F9D16.30"
              /number=2
```

```
Query Match       5.7%;    Score 94.2;  DB 7;  Length 119430;
Best Local Similarity 56.5%;   Pred. No. 1.9e-05;
Matches 196;  Conservative   0;  Mismatches 148;  Indels   3;  Gaps   1;

Qy    440 agcgccgtagcagagcacttgtcacagggcgtgccctacaagctatcggccgacgacgtc 499
           || |||||||||||||||| |  | ||||| | |  |||||||| | ||| || |||||
Db  29109 AGTGCCGTAGCAGAGTATCTAAACCAAGGTCTTCCAAAGAAGCTTACCGCAGATGACGTG 29168
```

us-09-026-400-1.rge